Place Recognition and Navigation in Virtual Environments

Dissertation
der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Diplom-Bioinformatiker
Marc Halfmann
aus Kandel

Tübingen
2015

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Prüfung:	11.5.2016
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Hanspeter A. Mallot
2. Berichterstatter:	Prof. Dr. Martin V. Butz

Zusammenfassung

Von der alltäglichen Vielfalt kognitiver Aufgaben, sind Ortserkennung und Navigation von großer Bedeutung. Wir müssen uns fortwährend in unserer Umgebung orientieren und Bewegungen oder Routen planen, um von einem Ort zum anderen zu kommen. Dieses Orientieren und Planen erfordert eine Reihe kognitiver Prozesse. Die in dieser Arbeit präsentierten Studien dienten dazu, diese verschiedenen Aspekte zu untersuchen.

Um uns zum Beispiel innerhalb eines Raumes zu bewegen, benötigen wir eine räumliche Repräsentation in unserem Arbeitsgedächtnis. Dieser so genannte Orts-Code basiert auf den räumlichen Sinnen, z.B. Sehen, und wird kontinuierlich aktualisiert, während wir uns durch unsere Umwelt bewegen. Was einen solchen Orts-Code ausmacht, d.h. wie Orte in unserem Gehirn gespeichert werden, ist weiterhin unklar. Ein möglicher Orts-Code ist das so genannte "*Spatial Image*", eine weniger bildhafte, abstraktere Repräsentation des Raumes, in der die Distanz und die Richtung zu umgebenden Objekten gespeichert wird. Die Frage ist, welche minimalen Umgebungsreize für den Aufbau dieser Repräsentation benötigt werden. In realen Umgebungen steht eine Vielzahl an Reizen zur Verfügung. Im Gegensatz dazu verwenden die hier präsentierten Experimente eine virtuelle Umgebung, in der die Umgebungsreize auf reine Tiefeninformation reduziert wurden. Die Umgebung wurde entweder auf einem Spiegelstereoskop oder einem Head-Mounted Display dargestellt. Versuchspersonen wurden in einer einfachen Ortserkennungsaufgabe getestet. Die Ergebnisse zeigen, dass der Aufbau eines "*Spatial Image*" rein auf Tiefeninformation basierend möglich ist, die Ortserkennung sich aber verbessert, wenn mehr Umgebungsreize verfügbar sind. Des weiteren zeigt sich, dass Ortserkennung in virtuellen Umgebungen von einem stärkeren Präsenz Gefühl profitiert.

Das Planen einer Route von einem Ort zum anderen erfordert räumliches Wissen über unsere Umwelt im Langzeitgedächtnis. Man nimmt an, dass Plätze in einer solchen Repräsentation in einer Graphen- oder Kartenähnlichen Struktur verknüpft sind. Beim Ablaufen einer Route, werden Teile davon ins Arbeitsgedächtnis transferiert. Dieser Prozess des räumlichen Abrufs kann durch eine Reihe von Faktoren beeinflusst sein, abhängig vom aktuellen räumlichen und aufgabenbezogenen Kontext. Das hier durchgeführte Experiment untersucht räumlichen Abruf während des Laufens einer Route in einer virtuellen Stadt. Versuchspersonen sollten Übersichtskarten bekannter Orte in Tübingen zeichnen. Die Ergebnisse zeigen, dass der räumliche Abruf für Orte die auf der Route liegen von der Anmarschrichtung abhängt. Im Gegensatz dazu wird der räumliche Abruf von Orten, die abseits der Route liegen, von der Luft- und der Kopfrichtung beeinflusst. Dies zeigt ebenso, dass derartige Experimente in immersiven virtuellen Realitäten durchgeführt werden können, und dass die Ergebnisse mit Studien in realen Umgebungen vergleichbar sind.

Summary

Among the manifold of cognitive tasks that we have to deal with in our daily life, place recognition and navigation are of great importance. We constantly have to orient ourselves within our environment and plan movements or routes to get from one place to another. In this orientation and planning procedure, different cognitive processes are involved. The studies presented in this thesis were aimed at investigating these different aspects.

To navigate, for example, within a room, we need a spatial representation in our working memory. This so-called place code is build up based on input from spatial senses, e.g., vision and is constantly updated as we move through our environment. It is still unclear what constitutes such a place code i.e. how are places actually stored in the brain. One possible place code is the so-called spatial image, a less picture-like and more abstract representation of space where distance and direction to surrounding objects are memorized. The question is, what are the minimum cues required to build up such a representation. In real world environments a large variety of cues are available. Here, in contrast, we present a series of experiments using a small-scale virtual environment, where the available cues were reduced to depth information only. The environment was presented either on a mirror stereoscope or a head-mounted display. Participants were tested in a simple return to cued place paradigm. Results show that building up a spatial image based on depth information only is possible but performance increases as more cues become available. Also place recognition in virtual environments benefits from a higher sense of presence.

The planning of a route from one place to another requires spatial knowledge of our environment stored in longterm memory. In such a representation, places are assumed to be connected to form a map- or graph-like structure. When traveling a route, parts of this structure are transferred into working memory. This process of spatial recall maybe influenced by a variety of factors depending on the current spatial and task-related context. Here an experiment was conducted to test spatial recall during route navigation in a virtual city. Participants were asked to draw sketch maps of well-known places in the city of Tübingen. Results show that for places that are lying along the route, spatial recall depends on the approach direction. In contrast, for places off the route there was an influence of aerial and head direction. The results also show that such experiments can be carried out using an immersive virtual reality setup, yielding results that are comparable to real world studies.

Acknowledgments

At this point I would like to thank all the people that helped and supported me in writing my thesis over the past years.

Foremost, I would like to thank my supervisor Prof. Dr. Hanspeter A. Mallot for supporting my work and for giving me the opportunity to work at the Cognitive Neuroscience group at the University of Tübingen. His guidance and advice made this thesis possible.

I am very grateful to Prof. Dr. Martin V. Butz for his willingness to review my work in a short time.

I would like to thank all my colleagues at the Cognitive Neuroscience group for their support and for many valuable discussions over the past years. Their feedback and suggestions were always a source of inspiration.

My thanks also goes to all the bachelor students that helped me collecting the data, namely Marie Admard, Viktoria Prozmann and Nina Walker for the kite-shaped room experiments and Nils Brehm for the Virtual Tübingen studies.

I would also like to thank Michaela Mohr for patiently proofreading my thesis.

My work was funded by the Bernstein Center for Cognitive Neuroscience, BMBF FKZ: 01GQ1002. I thank all the people in the Bernstein network that made this work possible and supported me over the years.

Finally I would to thank all my friends and family who supported me throughout my studies and my PhD thesis. My special thanks go to my dear parents who were always there for me and encouraged me on my way.

Contents

Chapter 1

Introduction

The topic of this thesis is the recognition of places and the navigational behavior of human participants in virtual environments. Here several studies are presented that examine basic mechanisms of spatial cognition as well as higher level abilities. All these studies were carried out in virtual environments of different scales, using either a stereoscopic setup of computer monitors or an immersive virtual reality system. The layout of this work is structured in five chapters as follows.

First, an introduction is given together with background knowledge and an overview of previous work in this field. Place recognition and navigation in humans are explained in general, and a more detailed illustration of human depth perception is given. As all experiments presented in this work were done in virtual environments, this topic is also covered in detail in the introduction.

The chapters 2-4 present the studies that were conducted within the scope of this work. Chapter two and three cover the experiments that were carried out in small scale environments, concerning basic mechanisms of place recognition. In particular the question is what are the basic cues that are used in the formation of place codes when a representation of a novel place is stored in longterm memory (LTM). In chapter four an experiment is presented that uses a large scale virtual environment. This experiment was aimed at answering the question how the representations mentioned above are transferred from longterm to working memory (WM) in a "real world" navigation task.

Finally chapter five presents several studies that were not conducted as part of this thesis but to which substantial work was contributed over the past years. This contribution includes designing and programming of experiments but also the modeling of three dimensional objects and environments used in the respective experiments.

1.1 Place Recognition and Navigation

Among all the different tasks and challenges we are dealing with in our daily life, place recognition and navigation play an important role. We are constantly facing the situation that we have to go to places or locations that we have either encountered before or which we are visiting for the first time. We are more or less easily able to recognize a known place when returning there even if the place looks slightly different due to changes at the place itself or just changes, e.g. in illumination due to different weather conditions. For example when revisiting the marketplace in a known city after years, one will still recognize the place although some of the store buildings may have changed.

To be able to achieve this, different steps are needed. First, the new place is learned and a representation has to be stored in longterm memory. Whenever we want to return to such a previously visited place, this representation is then transferred from longterm into working memory. Finally, to find our way from our current location to the desired target place, there has be a connection between those places within our longterm memory that allows us to plan a route.

1.1.1 Representations of Places

When encountering a novel place the question is how this place will be represented in our memory, i.e. what are the so-called place codes. There are several cues or characteristics of a place that could serve as such a code. These include basic ones like snapshot images but also higher level codes such as landmark configurations or spatial layouts.

The first idea one could think of when challenged with the task to return to a previously visited place, is to use a simple snapshot matching algorithm. Mallot et al. defined a snapshot as "the 'local position information', i.e. the manifold of all sensor readings as a function of observer position and orientation at one particular point (a raw image)" (Mallot and Gillner, 2000). The basic idea is to take a snapshot at the place that is to be memorized and store this snapshot in longterm memory. When returning to this place the stored snapshot is perpetually compared to the current view of the environment and the place is recognized when the difference between these two images falls beneath a certain threshold. For example, Cartwright and Collett showed this searching strategy for honey bees (Cartwright and Collett, 1983). They trained bees to find a food source with its location defined either by a single cylindrical landmark or by an array of landmarks. The food source was then removed and the bees were tested with either landmarks of different size or at different distances from the original location of the food source. The results show that "the bees then always search where the compass bearings of the landmarks on its retina were the same as they had been when it was stationed at the food source".

Therefor it can be concluded that honey bees indeed use a snapshot-based strategy for place recognition instead of the distances to the landmarks. However, in contrast to insect research where snapshot navigation is well established, only little is known about the usage of snapshots in human place recognition. That "human memory for places can make use of a snapshot algorithm" was demonstrated by Gillner et al. (2008). They tested subjects in a return-to-cued-location task that was carried out in a virtual reality setup. The task was to learn a goal location in a circular or a square simulated room and then find back to this goal from another location in the room as accurately as possible. The walls of the room were covered with a continuous color gradient spanning the entire color cycle with no other landmarks placed in the room. The experimental results were compared to the predictions of four different models: image difference model (snapshot), egocentric depth model (distance to wall), closest wall segment, boundary vector cell model. The only predictions that turned out to be consistent with the observed results are the ones from the image based model. They therefor concluded, "that humans are able to approach and recognize places based on raw image information void of localized cues, landmarks, or objects" and that "this performance is similar to visual homing behavior described for various insects".

Another possible way to memorize a place is by using landmarks. Although the concept of a landmark is widely used, it is not clearly defined what constitutes a landmark and even the differentiation between landmarks and snapshots is sometimes vague. In a more geographic sense a landmark can be defined as "An object or feature of a landscape or town that is easily seen and recognized from a distance, especially one that enables someone to establish their location." (Oxford Dictionaries, *http://www.oxforddictionaries.com*). Therefor a landmark has to be salient and distinguishable from its surroundings to allow us to use it for orientation and navigation. Also a landmark object has to be permanent to make it possible to recognize a place when returning. As defined by Siegel and White (1975), not only single objects but also a configuration of objects can make up a landmark. For human navigation, landmarks are of great importance and are used in many ways. For example, we can use a distant visible building, e.g. a church tower, as a beacon to return to a place. Likewise a configuration of such distant landmarks may be used to memorize a place with relation to this configuration via triangulation. Another possible use is in route navigation. Here, landmarks that are lying along a route may serve as decision points by being associated with an appropriate movement. Finally, landmarks may be used for navigation without being directly linked to the desired target location but serving as a kind of compass direction. A distant mountain, for example, can be used to keep a constant heading direction during navigation.

In animal place recognition, although landmarks seem to be a reliable cue for homing, they sometimes are being ignored. Cheng showed that rats seem to rely more on geometrical cues than on landmark information or visual features in general (Cheng, 1986). When rats were trained to find food in the corner of a rectangular box, they searched in the correct corner but also made a significant amount of systematic errors searching in the opposite corner, i.e. the corner with the same geometric properties. These errors appeared despite the fact that one wall had a different color and there were panels placed in each corner that differed in visual, tactile and olfactory characteristics and which could have served as landmarks. Cheng therefor concluded that rats use geometric relations and the shape of the environment for navigation, and that this was done by using what he called a "purely geometric module". Although this geometric module was examined and confirmed by other researchers in the following years, more recent research hast cast doubt on this concept. For example, Graham could show that rats performed badly in a task where visual cues were confusing and geometrical cues would have clearly helped finding the goal (Graham et al., 2006). They trained rats in a kite shaped water maze with two adjacent black walls and two adjacent white walls forming a right angle. The platform that the rats had to reach was always located in the same right angle corner but the colors of the walls were flipped from trial to trial. Therefor rats could not rely on visual cues but they should be able to learn the geometrical relations of the corners using the geometric module. The fact that they failed to do so questions the concept of a geometric module that is impervious to non-geometric information. Referring to this work, Cheung et al. (2008) was able to show that the findings of Graham et al. could be explained by applying a simple snapshot matching algorithm. They ran computer simulations using the same layout and visual properties. Within this simulated environment, a panoramic snapshot was taken at the goal location and then compared to snapshots taken at points of a regular grid covering the whole room by taking the mean squared pixel distance (MSPD). They plotted the MSPD for each grid point showing minima (global and local) of the image difference function. A simple homing algorithm that always choses the next grid point which minimizes the MSPD until it reaches a minimum was used to calculate catchment areas for all minima. A simulation of search trials based on these catchment areas and using the MSPD yielded results comparable to those of Graham et al. again questioning the postulations of Cheng. In a review from 2008 (Cheng, 2008) Cheng also argued, referring to the studies mentioned above, against the concept of a purely geometric module which is not affected by other e.g. visual cues.

The disadvantage of using a snapshot matching algorithm in human place recognition is clearly the lack of a panoramic view of the environment. Such a panoramic snapshot, as for example used in robot navigation (Franz et al., 1997) or the computer simulations mentioned above, assumes that the entire surrounding is visible at all time during navigation. Since in human vision the horizontal field of view is close to 180°, there must be some mechanism for place recognition to take into account perceptual information, e.g. visual cues that we have recently perceived, but which are no longer within our current view. One possibility to deal with this requirement is the concept of a so-called spatial image as proposed by Loomis (2013). It is defined as a "a spatial representation that is relatively short-lived and, as such, resides within working memory. It plays an important role in the control of action in three-dimensional (3-D) space when task-relevant perceptual information is no longer present". A simple example given for this spatial image is to take three or four objects within ones current environment, memorize their position and walk forward with eyes closed. When asked to stop and to turn or point toward one of the memorized objects, this gives us an idea of how well we spatially updated the object's location relative to our own position. The spatial image is therefor a representation of the locations, i.e. the relative direction and distance of objects within our environment with respect to our own location and alignment. According to Loomis, the spatial image has specific characteristics that distinguishes it from the visual image (snapshot) as summarized in table 1.1. Another important aspect of the spatial image is that it is

visual image	**spatial image**
retains properties of visual percept (color, texture, shape)	external to observer, retains distance and direction information
picture-like, i.e. no parallax and fixed within imaginal space	spatially updates during observer motion, relative parallax
experienced in the anterior half of surrounding space	can exist in all directions

Table 1.1: Different characteristics of the visual image and the spatial image as defined by Loomis

"multisensory in origin", i.e. it can be formed in spatial working memory from input of all spatial senses (visual, auditory, haptic) but also from spatial language. Furthermore, the instantiation of the spatial image is not restricted to current sensory input, but can also be formed by recall of spatial information from longterm memory. Figure 1.1 shows a diagram which illustrates this concept.

1.1.2 Building a Cognitive Map

Concepts like snapshot matching or a spatial image enable us to navigate within our current environment and complete, for example, simple homing tasks. However, when we want to navigate from our present location to another place, maybe one that is further away as, for example, in another part of a larger city, we need to plan a route to get there. Therefore we not only need a representation of the places which resides in our longterm memory, but also some sort of connection between those places that tells us how to travel

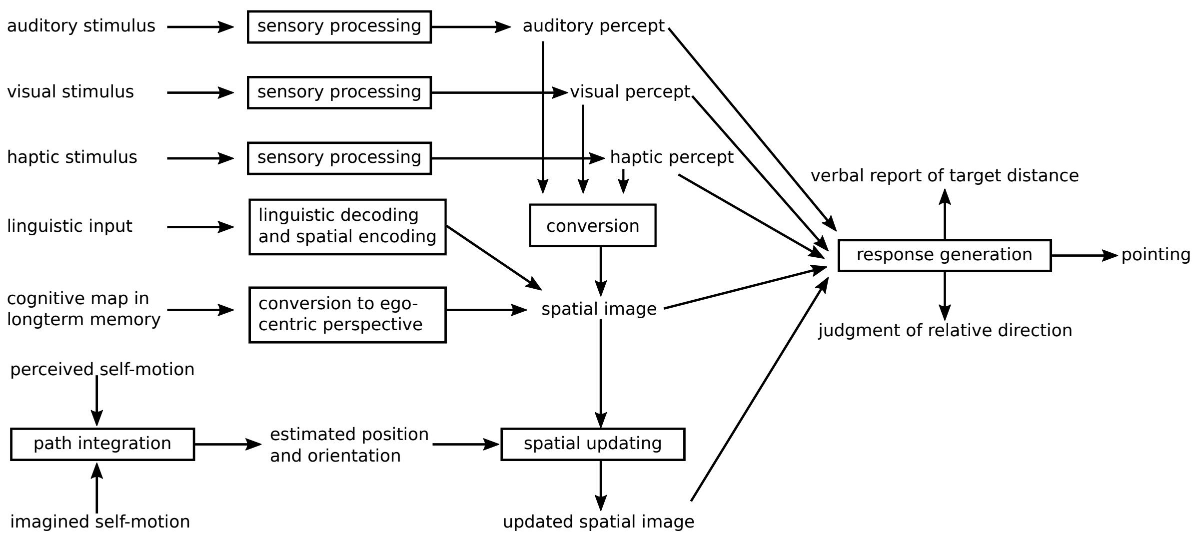
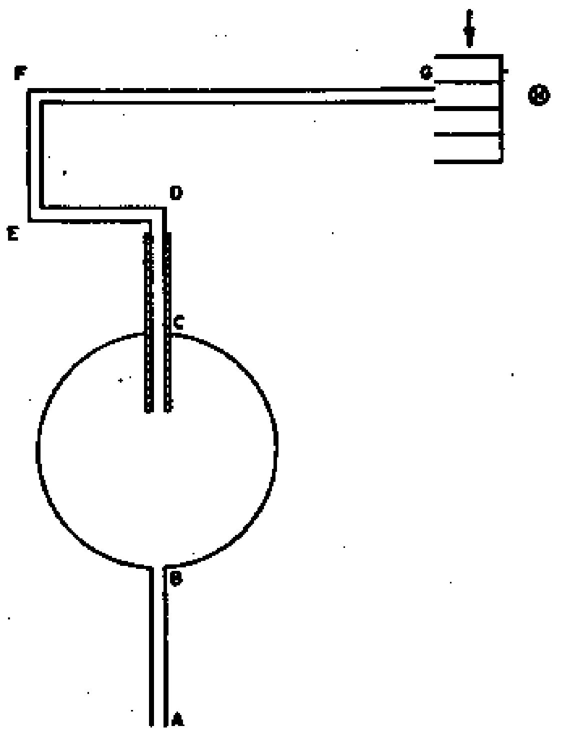


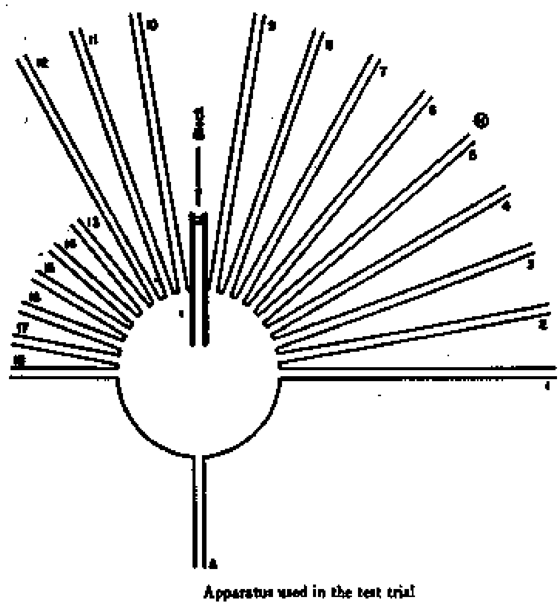
Figure 1.1: Block diagram of the spatial image (redrawn from Loomis (2013)). The spatial image receives input from all spatial senses, language and longterm memory. This representation is then updated either via perceived or imagined self-motion. Thus the spatial image allows to keep track of distance and directions of objects during observer motion.

from one to another. Such a connected representation is often referred to as a cognitive map.

The term cognitive map was originally introduced by Tolman (1948) in his pioneering work. Tolman presented several studies that lead him to this postulation, two of which will be described here in detail. In the first experiment three groups of rats learned a six-unit alley maze. All groups did one trial a day and were rewarded with food in the goal box of the maze. Whereas group one (control group) found food from the first day, group two found food from day seven and group three from the third day. Neither group two nor group three did learn the route through the maze to the goal box as long as they were not rewarded. When reward was given, the error rate immediately dropped rapidly the next day. Tolman concluded that the rats in group two and three had learned about the structure of the maze in the unrewarded trials. As soon as the food was present, the rats were able to make use of this knowledge gained during previous trials and showed an increased learning rate. In the second experiment, in the original paper referred to as "Spatial Orientation Experiments", rats learned an elevated maze as shown in figure 1.2a. Rats ran through this maze and found food placed in the goal box. They were trained four nights with three trials per night. After the training-phase the maze was altered so that now a series of radiating paths lead away from the table as illustrated in figure 1.2b. When placed in the changed maze rats first ran in the originally correct path, which was now blocked, and returned to the table. They then explored for a short distance the different radiating paths and finally chose one of them to run all the way to the end. Results showed that the path that was most frequently chosen by all rats was the one



(a) maze



(b) modified maze

Figure 1.2: Maze used by Tolman (1948) in the "Spatial Orientation Experiments". (a) Original layout of the maze. Rats started at A, ran across a circular table and entered the path leading towards the food box in the upper right. (b) Altered maze with the original path blocked and radiating paths leading away from the table.

that lead to a point close to were the food was originally located. Apparently, rats not only learned to run down the path that lead to the food location, but they also learned to chose a path leading in the direction of the food. Tolman therefor reasoned that the rats "acquired not merely a strip-map to the effect that the original specifically trained-on path led to food but, rather, a wider comprehensive map to the effect that food was located in such and such a direction in the room".

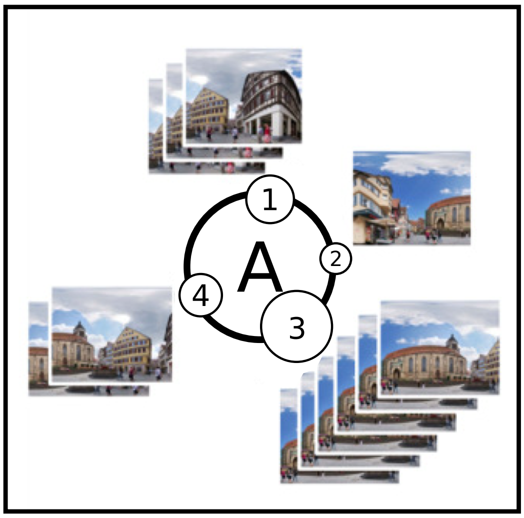
The neural correlates for these cognitive maps have been proven for rats. O'Keefe and Dostrovsky (1971) discovered cells in the rat hippocampus that fire whenever the animal is at a certain location within its environment. These locations are referred to as place fields. Evidence for place cells in humans have been shown by Ekstrom et al. (2003), by recording from neurons in the medial temporal and frontal lobe of patients with pharmacologically intractable epilepsy. While navigating a virtual environment, these cells exhibited a similar pattern of activity, firing whenever patients encountered a certain place in the environment. Another type of cells, so-called grid cells, were discovered by Hafting et al. (2005). These cells are "activated whenever the animal's position coincides with any vertex of a regular grid of equilateral triangles spanning the surface of the environment. Grids of neighbouring cells share a common orientation and spacing, but their vertex locations (their phases) differ". The existence of grid cells in the human brain is still unknown. Doeller and Burgess (2008) have found evidence for grid-cell-like representations in humans in a fMRI study with participants navigating in a virtual environment. Besides place cells and grid cells there is a third type called boundary vector cells (BVC) or border cells. A BVC model was proposed by Barry et al. (2006). In this model, place cells receive input via a feed forward connection from cells, the BVC that fire whenever the animal is close to an extended barrier (walls, large objects). Evidence for the existence of such cells in the subiculum of rats was reported by Lever et al. (2009). All these cells together are assumed to be the neural basis underlying cognitive maps.

Besides the discovery of such cognitive maps in rats, the existence of such a map-like spatial knowledge is well established in human research as well. However, it is still unclear how such a representation is stored in our longterm memory. Often the term cognitive maps is used in a more metaphorical way, rather than assuming a real map-like structure in the brain. Kaplan (1973) for example has stated that it is "far from a cartographer's map, however. It is schematic, sketchy, incomplete, distorted and otherwise simplified and idiosyncratic". Another distinction that arises when speaking of cognitive maps, is that between an egocentric and allocentric frame of reference. Egocentric means that the reference frame of our spatial representation is self-centered and changes as we move through our environment. As an example, one might say "the tree is to the left of me" and after turning around say "the tree is now to the right of me". The term allocentric however is less clearly defined. While common definitions are speaking of a world-based frame of reference, independent of one's current location, it remains unclear what this world based reference is. Röhrich et al. (2014) defined "allocentric memory as one that does not change as the observer moves", thus not referring to coordinate systems or global anchor points. Generally speaking, map knowledge and spatial longterm memory are assumed to be allocentric whereas spatial representations in working memory are egocentric.

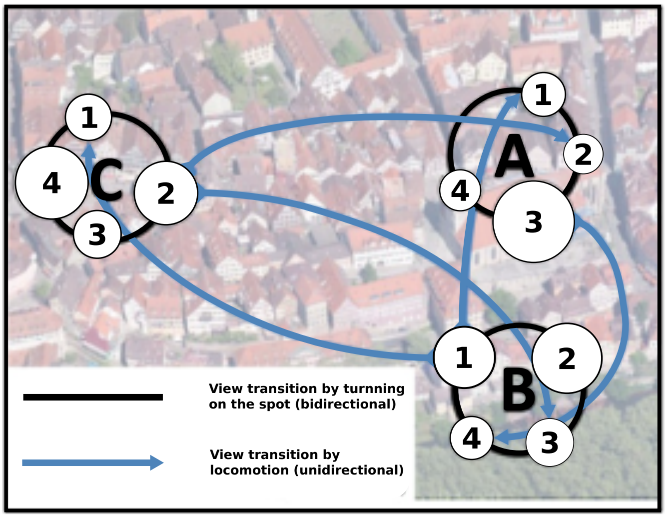
1.1.3 From Longterm to Working Memory

To be able to use this map-like spatial knowledge to plan routes and to navigate to previously visited places, we need to transfer this knowledge from longterm to working memory. The most popular working memory model was originally proposed by Baddeley (Baddeley and Hitch, 1974; Baddeley, 1986) and was later refined (Baddeley, 2000). In the current version the model consists of four components, the visuospatial sketchpad, the phonological loop, the episodic buffer and the central executive. As the names already imply, the visuospatial sketchpad temporarily stores and manipulates visual and spatial information, whereas the phonological loop is responsible for language information. The episodic buffer is a combined memory that can store visual and phonological information as short episodes. Superior and connected to all three of these components is the central executive. Its function is to connect to longterm memory and to distribute attention between the three other components. Important for the work presented here are of course the content of visuospatial sketchpad and the episodic buffer.

The question is, how is spatial knowledge transferred from longterm to working memory? In a study from 2012 by Basten, Meilinger, and Mallot (Basten et al., 2012) participants at the canteen of the University of Tübingen drew sketch maps of a well-known distant place in the historic city. The authors found that when participants were asked to imagine walking across this place, prior to drawing the maps, "drawings in the respective viewing direction became significantly more frequent". In a follow up study by Röhrich, Hardiess and Mallot (Röhrich et al., 2014) pedestrians in the historic city of Tübingen were stopped at locations around two central places and asked to draw sketch maps of the respective place. Results showed that map orientations depended on target and interview locations, i.e. drawings were influenced by the approach direction when walking from the respective current location towards the target place. The authors proposed a view-based representation of spatial knowledge in LTM as developed in (Schölkopf and Mallot, 1995). In this



(a) single place representation



(b) connected view-graph

Figure 1.3: Diagram of the view-graph model as proposed by Röhrich et al. (a) Representation of one single place. Multiple views of one place connected by a circle representing bidirectional connections of views by turning. Size of circles representing the frequency of this view. (b) Full view-graph of three places depicting transitions from one place to another by linking views. E.g., when standing at place C and seeing view C2, then walking to place B will end up in seeing view B3.

concept, referred to as a view-graph, places are stored as views where multiple views may exist for one place as shown in figure 1.3a. Views may be overlapping and anisotropically distributed regarding the viewing direction. Also, salient views, for example a prominent building, might be represented by multiple instances of this view or largely overlapping adjacent views. Connections between places are then stored by linking two views from adjacent places with a respective action label such as "follow the street from here". For example, when standing at place C in figure 1.3b and seeing view C2 then walking to place B will lead to perceiving view B3. Following this concept, the authors then suggest the spatial working memory as being a "subgraph of the full view-graph, consisting of the current view corresponding to the observer's current position and orientation, and the views reachable from this current view in a small number of steps". Therefor, when participants imagined the target place, they recalled the view connected to their view at their current location. Thus the spatial recall is depending on the interview location.

As stated above, the spatial image can also be based on inputs from longterm memory. Rieser for example (Rieser et al., 1994) asked young children to imagine their classroom while being at home. Then, in a walking condition, they had to walk around and imagine walking a path in their classroom that matched their current motion. Compared to an imagine-only condition, children in the walking condition were able to spatially update their mental representation. In a study from 2012, (Giudice et al., 2012) participants learned a set of three objects (LTM) and after a short disruption another set of three objects (WM). Participants were able to recall these LTM objects and integrate them into their current spatial image together with the current WM objects.

1.2 Perception of Depth

When we look around, we do not perceive our surrounding world as being flat, but rather three-dimensional and hierarchically structured in continuous depth planes. Perceiving depth allows us to judge distances to objects within our environment and is therefor essential for many common tasks. This includes fairly basic tasks, as for example grasping objects, but also more complex tasks like navigation and place recognition.

There are several visual cues that we can use to perceive depth. On the one hand there are monocular cues such as occlusion or motion parallax, on the other hand there is stereo disparity which requires of course binocular vision. An overview of different monocular cues is given in the following section, for a more detailed review see Vishton and P.M. (1995). Stereo vision is of particular importance for the presented studies and will therefor be explained in detail in 1.2.2.

1.2.1 Monocular Cues for Depth Perception

Texture Gradient

When looking at the structure of a surface, for example the bricks of a wall, then the elements of this structure appear smaller, the farther away they are from the observer.

Therefor the bricks are perceived as being more densely packed on the surface. This so-called texture gradient gives an impression of depth.

Aerial Perspective

Aerial perspective refers to the influence of the atmosphere on the perception of distant objects. With increasing distance to the observer, the contrast between objects decreases and the color changes. For example, when looking at a mountain range at the horizon, the farther away the mountains are, the more bluish they appear.

Motion Parallax

Parallax is defined as "the apparent displacement of an observed object due to a change in the position of the observer" (*http://dictionary.reference.com/browse/parallax*). The greater the distance from the observer to the object, the smaller the amount of displacement. That means that as we move through our environment, objects at farther distances appear to move slower than objects at closer distances. We know this phenomenon, for instance, when traveling by train and looking out of the window. Houses and other objects that are close to the railway are passing by faster than for examples the mountains at the horizon. Motion parallax is a very strong cue to perceive depth but needs movement either of the observer or the perceived image.

Relative Size and Size Constancy

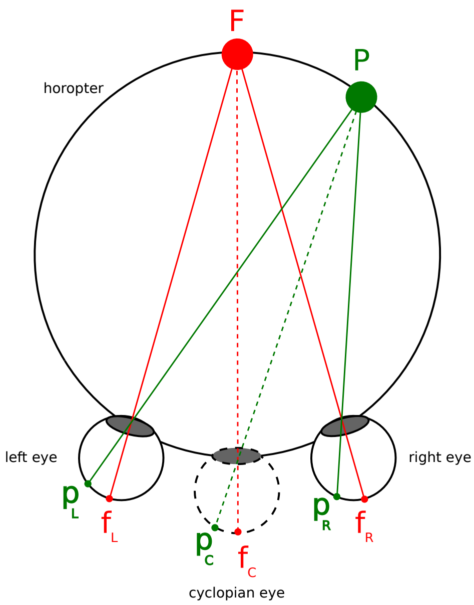
Relative size means that if two objects that are known to be of the same size, then their relative size gives information about the relative depth of these objects. That means if the two objects are at different distances from the observer, then the one that subtends a larger visual angle on the retina must be closer. Size constancy refers to the fact that if the absolute size of an object is known, than the size of the retinal image gives information about the object's distance to the observer, since the size of the retinal image decreases with distance.

Occlusion

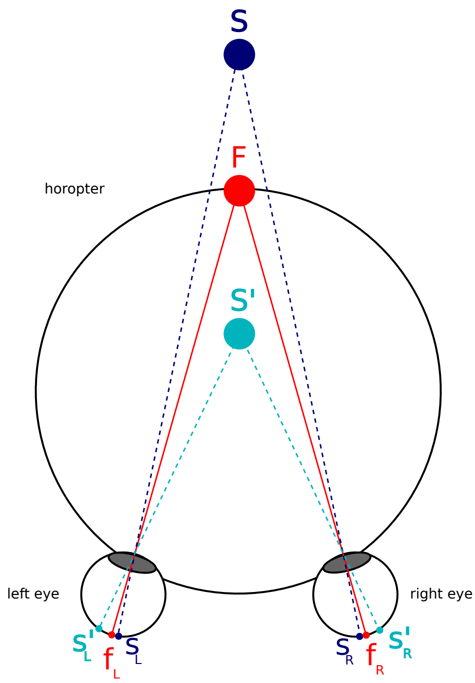
One very obvious cue for perceiving depth is of course occlusion. Occlusion occurs when one object partially hides behind another object that is at a closer distance to the observer. This allows us to get a depth ordering of different objects but does not give information about the absolute distance of an object.

1.2.2 Stereo Vision

The term stereo vision or stereopsis refers to the ability to perceive depth or 3-dimensionality based on binocular vision. The basic requirement for stereopsis is therefor having two eyes, located at different lateral positions on the head. Due to this different lateral positions, each eye sees the same object from a slightly different angle, leading to slightly different images projected onto the retina. The difference in position between those two images is called horizontal disparity. Assume, an object F is fixated binocularly, then this object



(a) binocular fixation



(b) horizontal disparity

Figure 1.4: Schematic depiction of the principle of binocular vision. (a) When fixating point F all other points, e.g., point P that lie on the horopter, are projected onto corresponding retinal areas. (b) Objects that lie in front or behind the fixation point are projected onto non-corresponding areas and have either crossed (S') or uncrossed (S) disparity. Disparity increases with distance from the horopter.

is projected onto corresponding areas f_L and f_R on the retina as shown in figure 1.4a. Objects for which the projected images lie in corresponding areas have a zero-disparity and are therefor seen as one single object. Now, when fixating an object, there are other points, e.g. point P in figure 1.4a, besides the fixation point that are projected onto corresponding areas as well. The entirety of all these points lie on the so-called horopter, a theoretical circle going through the fixation point and the focal point of the eyes. For objects that are either closer or farther away than the fixation point, the images are projected onto non-corresponding retinal areas, creating horizontal disparity. Horizontal disparity increases with distance between the object and the horopter. As can be seen from figure 1.4b objects that are closer to the observer than the fixation point (S') appear right of F in the left eye and left of F in the right eye. This is referred to as crossed disparity. Likewise objects that are farther away than the fixation point (S) are seen right of F in the right eye and left of F in the left eye (uncrossed disparity). Horizontal disparity therefor allows us to judge the distance to objects and is especially important within "grasping space".

Experiments have shown that not only fixated objects are fused into one single image but also objects that are within a small distance in front of or behind the horopter. This range is called panum area, objects outside of this area are seen as double images. Objects within the panum area are perceived as being seen by one single so-called "cyclopian" eye, a theoretical eye located between the left and right eye (see f_C and p_C in figure 1.4a).

All methods that are currently used to present 3-dimensional content to the viewer, e.g. in cinemas or computer simulations are based on this principle. A variety of examples for these methods are explained below in 1.3.2.

1.3 Virtual Environments

All the experiments presented in this thesis were carried out in virtual environments that were presented either on a single-mirror stereoscope or a head-mounted display. The advantages of such computer simulations have been explained in detail by Bülthoff and Christou (2000). The following section will give a brief overview of these advantages as well as the drawbacks of virtual environments.

1.3.1 Basics of Virtual Environments

Earlier studies in spatial cognition were done in real world environments which were often custom built and therefor costly and time-consuming to produce. Also the stimulus conditions in reality are not always easily controllable. On the other hand, the increase in computing capacity of modern CPUs and graphic cards has led to a high level of realism in virtual environments. At the same time, the hardware necessary to allow such realism has become widely available at reasonable prices, making the setup affordable for many laboratories. Thus the usage of virtual environments has become a well established tool in investigating spatial cognition.

In their work from 2000, Bülthoff and Christou used the terms virtual environment, or virtual environment simulations in their case, and virtual reality to relate to the same thing. However, they noted that many people used the term virtual environment, "having realized that definitions of what ”reality” actually consists of are somewhat problematic”. Also, the term virtual reality is broader defined, meaning that it offers the user an experience that gives a stronger feeling of actually being in this computer simulated world. Especially nowadays, with the emergence of many affordable HMDs like the Oculus Rift, the HTC Vive or even low cost solutions like Google Cardboard, the term virtual reality has become more and more related to these kind of experiences. They allow users to look around and interact with the environment in a way that goes beyond the possibilities of ”classical” environments presented on computer monitors. In the following we will stick to this distinction between those two terms.

The benefits of using virtual environments are manifold. First, they provide the possibility to control the environment and thus the presented cues and stimuli in nearly every aspect. For example computer simulations allow a precise control of the lighting conditions, whereas this parameter is more difficult to control in real world settings. It is also possible to easily change stimulus parameters, such as the size of landmarks, from trial to trial. Second, the number and type of environments is practically unlimited, allowing the creation of novel environments, especially such environments that could not be realized as real world setups.

However, the usage of virtual environments and virtual reality comes with a number of drawbacks. The first is that the degree of realism is still limited. It is, due to technical limitations, not possible to implement a computer simulation that creates exactly the same perception as the real world. The key concepts here are presence and immersion. Slater et al. (1996) illustrated the difference between the two terms. According to them immersion is the entirety of what the simulation provides, i.e. what input it offers to the user. Presence on the other hand is the sense of actually being in the Virtual Environment. Thus a higher immersion leads to a greater presence. On the visual part of the simulation, computer generated environments are getting more and more realistic nowadays while still being far from photo-realistic. The bigger problem is to incorporate other senses such as proprioception. This is particularly a problem when viewing scenes on a normal computer screen or the stereoscope and can be overcome partially by using head-mounted displays with included head tracking. The possibility to look around freely in a virtual environment greatly increases ones presence. This can be increased even further by not only tracking the head position, but also the body, and transferring participants real world translation into virtual reality. The problem here of course is that this limits the size of the virtual environment to the size of the real world room where the experiment takes place.

The second problem comes with the potential impact of motion- or simulator-sickness. This is due to the fact that when experiencing virtual reality simulations, the visually perceived movement contradicts the movement perceived by the vestibular system. This is not a big problem for environments presented on a computer screen, at least for younger participants who are more used to deal with such simulations due to playing video games. However, this becomes a bigger problem when using HMDs, which provide a greater immersion and thus a bigger discrepancy between visual and vestibular input.

1.3.2 Displaying Virtual Environments

Virtual Environments can be presented on various devices such as single computer monitors or projection screens and HMDs for a more immersive experience. Whereas environments presented on a single monitor contain all the features mentioned in 1.2 for monocular depth perception, they do not provide stereoscopic depth cues. That means they appear rather flat to the observer than being perceived as 3-dimensional. As the experiments presented in chapter 2 and 3 were aimed at investigating the role of stereoscopic depth cues in place recognition, we needed a setup that provided participants with a 3-dimensional view of the environment. To achieve such a stereoscopic presentation of virtual environments, a large variety of methods are available which will be explained in this section. All these methods have in common that the scene is rendered twice each frame, creating two different partial images, one for each eye. The trick is then to ensure that each eye only sees the respective image.

Anaglyph 3D

This is probably the most well-known method. Here the two partial images are displayed overlayed and in complementary colors (e.g. red and green). Images are then viewed through colored glasses with corresponding colors (one for each eye), so that the red filter cancels out the red part of the image (green is seen as black) and the green filter cancels out the green part. This has the effect that both eyes see a different image, leading to a three-dimensional image in the brain. It is obvious that this method has the disadvantage of loosing color information in the image.

Shutter-3D-Systems

These systems use special glasses that can electronically cover the left and the right eye alternately. The computer monitor then switches synchronously between the left and right partial image, so that each eye always only sees the designated image. The disadvantage of this solution is the reduction of the number of frames per second. This can be overcome by using high frame rates (120Hz) that in turn require faster hardware to run.

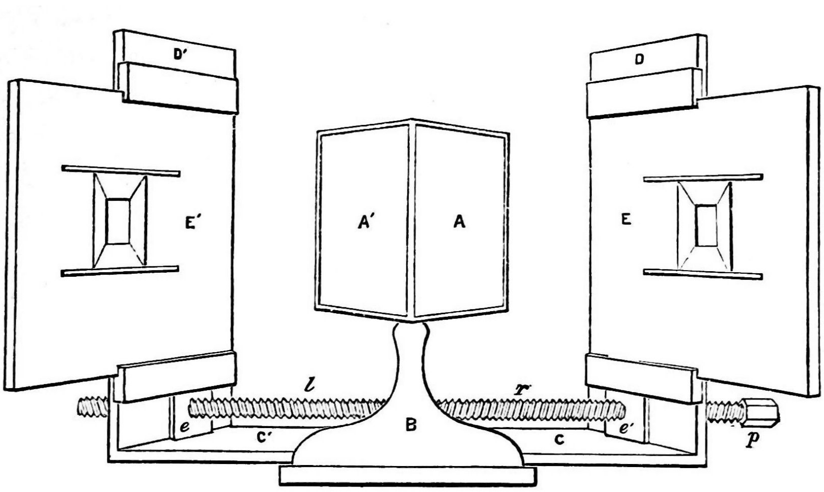


Figure 1.5: Sketch Drawing of the mirror stereoscope originally constructed by Sir Charles Wheatstone (Le Conte Stevens (1882)). The images for the left and right eyes are placed in the panels E and E' and are viewed by the observer via mirror A and A' respectively.

Polarization

Polarization is the method that is currently used in cinemas or in home television sets to present 3D movies. Here the two partial images are displayed using orthogonally polarized light. The observer views the movie through glasses with the respective polarization filter for each eye. Thus each eye only sees the image with the appropriate polarization. One disadvantage of this method is that when used on a screen (TV or computer), both images have to be shown simultaneously, which reduces spatial resolution. Also, the vertical viewing angle is narrower as for example compared to shutter-3D-systems.

Mirror Stereoscope

Mirror stereoscopes are an old setup for viewing images in stereoscopic 3D. The stereoscope was originally invented by Sir Charles Wheatstone in 1833, a sketch drawing of the original setup is depicted in figure 1.5. The original stereoscope contained two mirrors that were arranged perpendicular to each other. Lateral to the mirrors an image could be placed on each side of the apparatus. The arrangement of the mirrors and the respective image was aligned in a way that an observer facing the edge of the two mirrors could see one image with each eye only. Thus when presenting two images, which showed a scene from a slightly different perspective as in human stereo vision, a 3-dimensional image was perceived. Wheatstone first used drawings of scenes, but with the emergence of photography at that time, stereoscopic imagery improved and became more popular.

The stereoscope we used in our experiments is a modification of the original Wheatstone design and was proposed by Kollin (2007). A detailed description of our setup will be given in 2.1.1. The basic idea is to replace the two mirrors of the original arrangement with one single mirror, so that one eye looks directly onto one monitor and the other eye sees the mirrored image (figure 2.1a).

Using such an apparatus for presenting 3D virtual environments has the advantage that it does not suffer from the same drawbacks of the methods mentioned above. That means it allows to present each eye with an partial image at full spatial resolution and without reducing the frame rate or the color information.

Head-mounted display

As already mentioned above, HMDs are currently becoming more and more popular. The basic principal is the same for all these devices. Two displays, or one display divided in two images, are placed in front of the observer's eyes at a very short distance of only a few centimeters. To account for this short distance and to present a sharp image, lenses are placed between the display and the eye. Thus each eye only sees the image presented on the respective display. HMDs provide a greater immersion by offering a wide vertical field of view, but often have to compromise regarding image resolution. Also, for a comfortable experience, the frame rate has to be high (90-120Hz), requiring fast hardware for rendering realistic environments.

1.4 Scientific Issue

As already stated above, the topic of this thesis is separated into two different questions regarding human place recognition. The first question is concerned with the representation of places and the cues that are used to form a place code and learn a novel environment. Still little is known about such place codes. There is evidence that humans make use of snapshot navigation or that we build up a spatial image of our vicinity as we move through our environment. Input for this spatial image can arise from all spatial senses as well as from spatial language. As already explained, humans make use of stereopsis to judge distances to objects and this depth estimate is especially important within our vicinity. Therefor, the question that is asked here is whether place recognition is possible when all available cues are reduced to depth information only. In other words, can we build up a spatial image solely based on depth information and use this representation for a simple homing task. Results from a sparse environment providing depth information only are compared to a rich, textured environment. If depth cues are sufficient for place recognition, results from the sparse environment should be comparable to those from the control environment.

The second question is concerned with the recall of spatial information and the transfer of this information from longterm to working memory. It is generally believed that we have some spatial representation of our environment within our brain. This spatial LTM may be a map-like representation or organized as a graph-like structure, storing places as views. The experiments carried out in the context of this thesis follow the studies by Basten et al. and Röhrich et al. Therefor the question asked can be split up into two subquestions. First, as the previous studies were carried out in real world environments the question is, can these studies be reproduced in a virtual reality setup. That means, do the virtual reality experiments yield comparable results and if not, what may be the reason for this? Second, in the original experiments it was unknown what route participants were walking or where they came from. Here in virtual reality we were able to specify the route and control for what participants perceived. Thus it was possible to compare the spatial recall from places that were lying on the route, i.e. they were task-relevant, and from places that were off the route. If the transfer of representations into working memory depends on what participants expect to see next, there should be a difference between places on and off the route.

Chapter 2

Place Recognition in a Stereoscopic Setup

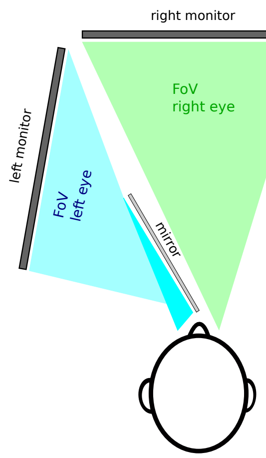
The studies presented here in chapter 2 were aimed at investigating basic mechanisms of human place recognition. More specifically, the question is if depth information alone is sufficient for building up a spatial representation to navigate within a small scale environment. The series of experiments described in this chapter were carried out in virtual environments presented on a single-mirror stereoscope. The setup of this stereoscope and the general methods were the same for all three experiments and are explained in detail in the following section.

2.1 General Methods

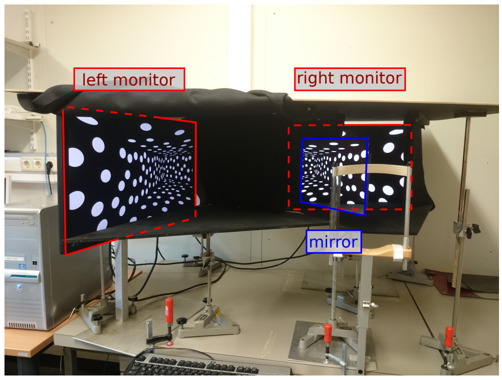
2.1.1 Setup

Apparatus

The stereoscope used in the following experiments was a single-mirror stereoscope constructed by Hannig (2012) and follows the design developed by Kollin and Hollander (Kollin, 2007). Instead of using two mirrors to present the two different images to each eye as in the original design by Charles Wheatstone, they suggested to use a single mirror placed between two monitors. A schematic sketch of our setup is shown in figure 2.1a. Participants were seated in front of the stereoscope with their head placed on a chin rest, looking with the right eye directly at the right monitor and with the left eye at the mirrored reflection of the left monitor. This setup ensured that each eye only saw the designated image and participants could perceive a 3D impression. The whole apparatus was covered with a visual cover and also the room was completely darkened during the whole



(a) stereoscope schematic



(b) stereoscope

Figure 2.1: Setup of the stereoscope. (a) Schematic of the setup. Two 27-inch monitors at an angle of 100° and the mirror in front of the observer. Field of view depicted for both eyes (green = right eye, blue = left eye). (b) Photo of the stereoscope showing the two monitors marked as red boxes and the mirror (blue box). Also shown is the visual cover and the chin rest.

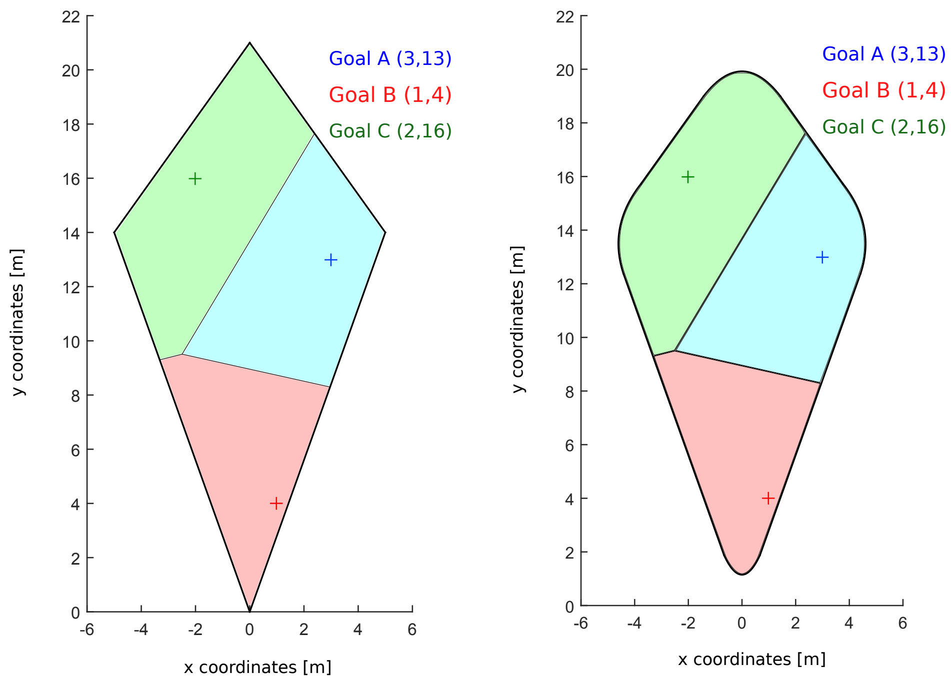
experiment, so that participants' view was not distracted by any external light source and they could only see the stimulus (figure 2.1b). The hardware used were two Samsung SyncMaster S27A850D 27-inch monitors, with an resolution of 2560 x 1440 pixels each, both connected two a PC with an Intel® Core™ i5-2400 @ 3.1 GHz with 8 GB working memory and an Nvidia® Geforce® GTX™ 570 GPU.

Virtual Environment

The environment used in all the experiments was a kite-shaped room with a size of 21 by 10 meters in which subjects were able to navigate freely using the mouse and keyboard. We defined three different goal locations within this room as depicted in figure 2.2a that were used for all experiments. Also shown is the voronoi partitioning of the room which is later used to define the errors.

There were two experimental conditions which differed in the way the stimulus was presented. In the first condition, the texture condition, the walls of the kite-shaped room were covered with a wallpaper of white dots (15 cm in diameter) on a black background (figure 2.3a). This condition was used as a control condition to test participants' place recognition in a "normal" virtual environment.

The second condition, the random-dot condition, was designed to test if place recognition can be based on depth information only. Therefor we created a stimulus where no other cues were available. This was done by using a random-dot pattern generator developed in Hannig (2012) to present our virtual environment. Whereas normally walls and objects in virtual environments are presented as textured polygons, in our stimulus all textures



(a) kite-shaped room (b) modified room with rounded corners

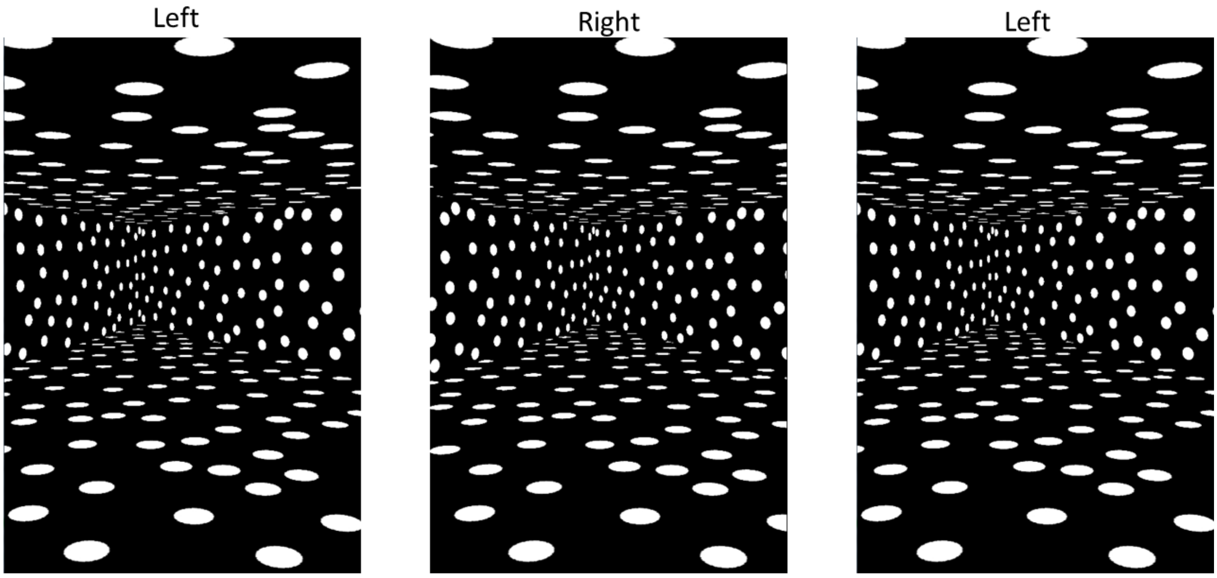
Figure 2.2: Layout of the kite-shaped room used for all experiments. Depicted are the three different goal locations with the respective voronoi segment. (a) Standard layout with sharp corners. (b) Modified layout with rounded corners.

were replaced by a pattern of dynamically generated dots. These dots were randomly placed on the surface of a wall, but were distributed in such a way that the whole wall was uniformly covered with dots. The number of concurrently visible dots was constant and their lifetime was randomly chosen between a minimum of 100 ms and a maximum of 200 ms. This means that each dot that died after its lifetime expired was immediately replaced by a newly generated dot at another position in the next frame. Figure 2.3b shows an example of participants' view in this condition. Both the image of the left and right eye are shown for crossed (right-left) and uncrossed (left-right) fixation. In this setup, the number of dots was set to 2000 and their size was set to 3 pixels.

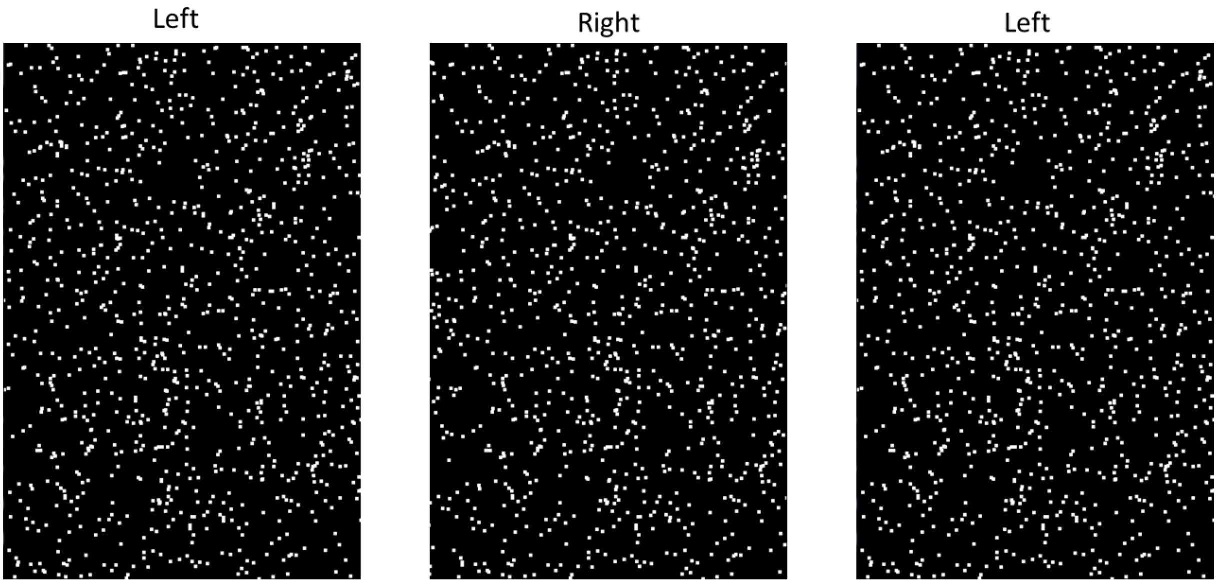
With the dots being visible only for a short time, participants were not able to perceive this pattern as a kind of texture and could not make use of snapshot matching or use other cues like texture gradient. Thus only depth information was available. In contrast, when a scene was presented normally using textured walls, participants were provided with a larger variety of visual cues for place recognition.

Software

For the programming and rendering of the experiment we used a modified version of OpenSceneGraph 3.0.1 (OSG), an open source 3D graphics toolkit for developing real-time



(a) tetxur



(b) random-dot

Figure 2.3: Example of participants' view in the virtual environment looking at the sharp corner. Images are arranged for free fusion, either crossed or uncrossed. (a) Texture condition. (b) Random-dot condition.

3D applications. It is written in C++ and uses OpenGL as a basic graphical framework. OSG already provides the possibility to display stereoscopic virtual environments supporting different methods, e.g. anaglyph or side by side stereoscopy (horizontal split). The second method displays two images that are shifted by a small amount on two different monitors. Because of the left image being mirrored in the setup used here, OSG's horizontal split method needed to be modified so that a mirrored view of the scene was displayed on the left monitor. This was simply done by changing the *computeLeftEyeProjectionImplementation* method's parameter *scale_x* to *-scale_x*.

2.1.2 Procedure

The procedure in general was the same for all the experiments except for the differences explained in 2.1.3. All participants did both conditions but on a different day, so that the duration of the experiment was not too long and participants did not lose their motivation. Prior to the random-dot condition participants had to do a short preliminary test as described below. The purpose of this preliminary test was to verify that participants could perceive a stereoscopic depth impression from looking at a scene presented as a random-dot stimulus. This also allowed us to get a measurement for participants' stereo vision and to see whether there was a correlation between participants' ability to perceive stereoscopic 3D and their performance in the main experiment. For all the experiments participants always started with the texture condition on the first day and did the preliminary test followed by the random-dot condition on the second day.

Before the start of the experiment participants were informed about the experiment itself, the collected data and the potential sickness caused by the computer simulation. They gave written consent that they had read and understood the terms of participation. Then they were given a printed summary of the experimental procedure that they could read through, but they were also instructed verbally by the experimenter and occurring questions were answered. As well they were shown a sketch of the room's layout so that they knew about its shape. After this they placed themselves comfortably in front of the stereoscope and the experiment began.

Preliminary Test

This preliminary test was done with the stereoscope using the same parameters (number of dots, dot lifetime etc.) as for the main experiment. Participants were presented two squares as random-dot stereograms on a black background and had to decide which one was closer to them by pressing the left or right mouse button. The squares were placed at a base distance of 10 meters away from the observer and the distance between the two squares was set to 0.25m, 1m and 5m respectively as shown in figure 2.4. Both

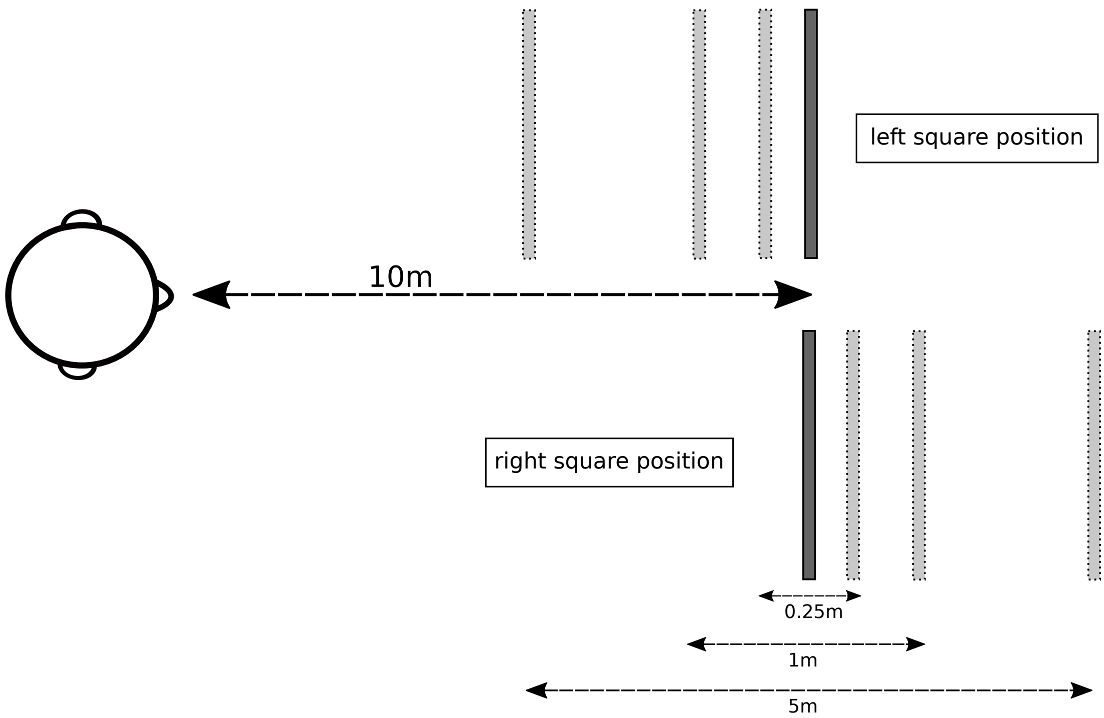


Figure 2.4: Setup of the stimulus in the preliminary stereo test, in this case the left square being closer to the observer. The dark gray boxes depict the base distance of 10 meters with no distance between the two squares (not used in the pre-test). The light gray boxes depict the inter-object distances of 0.25, 1 and 5 meters as used in the pre-test.

squares were adjusted in their side length so that they appeared to be of equal size to the observer. Each inter-object distance was presented 8 times with 4 times the left and 4 times the right square being closer to the observer, making up for a total of 24 trials in this preliminary test. Participants were not able to move within this setup so no distance information from motion parallax was available and they could only make their judgment based on disparity.

Main Experiment

For both experimental conditions (textured and random-dot) the procedure was as follows. At the beginning of each trial participants were shown a start screen with white text on black background and were requested to start the trial by pressing the space key.

After pressing the space key, the learning-phase began and participants were teleported to one of the three possible goal positions depicted in figure 2.2. During this phase of the trial they were able to turn around in the virtual environment by using the mouse and they were also able to move within a diameter of 1m around the goal position by pressing the W, A, S and D keys on the keyboard. The S and W keys were marked with a patch of felt to make it easier for participants to use the keys in the darkened room. Participants were instructed to look around in order to determine their current position within the virtual environment and to memorize this position as accurate as possible. When they felt confident that they had learned their position, they confirmed this by pressing the space key and then the test-phase began.

At the beginning of the test-phase participants were teleported to another location within the room, the start position for this trial, which was always one of the other two defined positions (figure 2.2). Participants were now able to move around freely within the virtual environment and their task was to navigate back to the goal position which they had previously learned in the learning-phase. When they thought that they had reached the goal position they confirmed their decision by pressing the space key. After pressing the space key they were rotated and moved passively towards the actual goal position to give them a feedback of how good they performed and where the real goal position was.

Then the start screen was shown again and participants were asked to start the next trial by pressing the space key. They were told that they should perform the learning and test-phase as accurate as possible, but that they also should try not to take too much time. There was always the possibility for them to take a short break between two trials when the start screen was displayed. Also, participants could abort the experiment at any time.

The sequence in which the three goal positions were assigned to the trials was not randomly chosen. Instead a pseudo-randomized sequence was predefined and the order of the goal positions was chosen in such a way that the goal position from the current trial was always the start position in the following trial.

Questionnaire

After they finished the main experiment, participants were asked to fill out a short questionnaire (A.1.2), rating the following questions on a scale from 1 (not at all) to 7 (high).

- Did you have fun doing the experiment?
- How was your motivation?
- How difficult did you find the experiment?
- Were there positions that were easier to find?

They were also asked to draw the goal locations that they could remember within a given sketch of the room's layout.

2.1.3 Experiments

Baseline

20 participants took part in the first experiment, all had normal or corrected to normal vision. The procedure was as described above. This experiment uses the basic setup with the kite-shaped room as shown in figure 2.2a and thus serves as a baseline. The findings from this experiment lead to the variations explained below.

Monocular

8 participants took part in this second experiment, all had normal or corrected to normal vision. In this experiment we tested the influence of motion parallax on depth perception and navigational performance. To test this we changed the setup described in 2.1.1 so that participants could no longer use stereoscopic vision to perceive depth information. This was simply done by covering participants' left eye with an eye patch, and allowing them only to view the right monitor. Motion parallax was now the only depth cue available in this setup. The procedure was as described in 2.1.2 with the exception that in this case no preliminary test was carried out prior to the random-dot condition, as without stereo vision there was no need to perform this test.

Rounded Corners

8 participants took part in this third experiment, all had normal or corrected to normal vision. The aim of this experiment was to test if participants could make use of some sort of landmark cues in the baseline experiment. This was because despite the fact that the limited life time of the random dots prevented participants from using a kind of snapshot representation, one could still argue that the sharp corners of the room might serve as some sort of landmark information.

To test this we changed the original layout of the room in such a way that there were now rounded corners instead of sharp corners. Figure figure 2.2b shows this modified layout. The goal positions remained the same as in the previous experiments. Also the procedure was the same as in the baseline experiment, including the preliminary test.

2.2 Results

2.2.1 Evaluation Methods

Preliminary Test

For the preliminary test we recorded participants' reaction time and answer (correct or false) for each trial. We then calculated the percentage of correct answers for each inter-object distance. Note that there was no measuring point for an inter-object distance of 0 meters. As the chance level for this task was 50%, we added this as a theoretical value for the 0 meter distance. The percentage of correct answers was then averaged over all participants and plotted against the respective distance (0, 0.25, 1 and 5 meters). To check for a correlation between participants' ability to perceive stereo vision and their performance in the main task we needed to define a measurement of their pre-test performance. For this purpose we calculated the theoretical inter-object distance where they answered 75% correctly, i.e.the threshold distance at which they could reliably determine which square was closer.

Main Task

For the main task we recorded a variety for variables for each trial. First, the time that participants needed to complete the trial was logged. This was split up into the time they spent looking around in the learning-phase (encoding time) and the time they needed in the test-phase to return to the goal location (search time). While the latter gives information about participants' navigational performance, the encoding time lets us draw conclusions about how fast participants perceived the room and their own position.

Second, we recorded the coordinates of the participants in the virtual environment for each frame. These data yielded the final location where participants confirmed their decision as well as the distance between the assumed goal location and the correct one. These decision points were then used to calculate the error distribution as described below. We also used the logged coordinates to plot the trajectories for each test-phase and to calculate the initial walking direction in which participants started their search. These two measurements give information about participants' search behavior.

To analyze participants' performance in the main task we defined two different error categories, qualitative errors and statistical errors. In a first attempt we defined a qualitative error by the distance between confirmed and correct goal location, considering every decision with a distance greater than 4 meters an error. This distance was chosen based on the histograms of the decision point's distribution which showed a clear clustering of confirmed positions within a radius of 4 meters around the respective goal. The problem with this definition was that it did not take into account the geometric layout of the room and the different characteristics of the goal locations. That means, for example, a decision point could still be counted as being correct although regarding the geometric relations it would be considered an error. Clearly for this task the geometric layout of the room and the relations between goal locations and the walls are an important cue and should therefor be taken into account when defining the error. Also choosing a threshold distance to define errors seems somewhat arbitrary and is difficult to justify.

We therefor used a more reasonable definition by applying the voronoi partition to the kite-shaped room. The voronoi partition divides a room into segments as determined by a set of points (centers) within this room. Each such segment is defined by one center and includes all points that, with respect to euclidean metric, are lying closer to this center than to any other center. In our case the centers are the goal positions and for each goal position all points that are lying closer to this than to any other goal position are defined as being correct. Figure 2.2 shows this partitioning. In the case of three center points the segments can easily be determined geometrically. First, the three points are connected by straight lines. Then, for each side of the resulting triangle, the perpendicular bisectors are drawn. Finally these lines are used to construct the edges of the voronoi segments. For easier error calculation, for each decision point the distances to all three goals was taken to decide whether it counted as an error or not. A subcategory of qualitative errors are rotational errors, i.e. trials when participants chose the geometrically opposite corner. Rotational errors were not analyzed explicitly but are sometimes helpful to explain participants' error rate.

The voronoi partition was also used to define the chance level. The size of each of the three segments was about $\frac{1}{3}$ of the whole room. Thus, for this definition, the simplifying assumption is made that a participant walking around randomly within the room will end up in one of the three segments with equal probability. Of course this depends on the starting point or segment, the search behavior and the time participants spent walking. However, in first approximation we set the chance level to $\frac{1}{3}$ or, in other words, the probability that participants made an error by chance was $\frac{2}{3}$.

The statistical error was defined as the distribution of the correct rated trials around the respective goal location. This error was quantified by the area of the error ellipses and the mean distance between the correct decision points and the respective goal location.

Besides participants' performance in finding the goal locations, we also looked at their searching behavior to see if there were differences between the two conditions. One possible aspect that can be examined here is their initial walking direction, i.e. the direction in which they started walking when they searched for the goal. As we did not record participants' viewing direction we had to calculate the initial walking direction manually.

To do this we took the location where they started and their coordinates after they had left a radius of 0.5m around the starting point. The Vector between these two points was then taken as the starting vector for this trial. For each trial we determined the difference between this starting vector and the vector pointing to the respective goal location. These differences were then pooled for all goals over all participants and the angular distribution was plotted as circular histograms. All circular histogram plots and the corresponding angular statistics were done using the CircStat toolbox for Matlab by P. Berens (Berens, 2009).

2.2.2 Experiment 1: Baseline

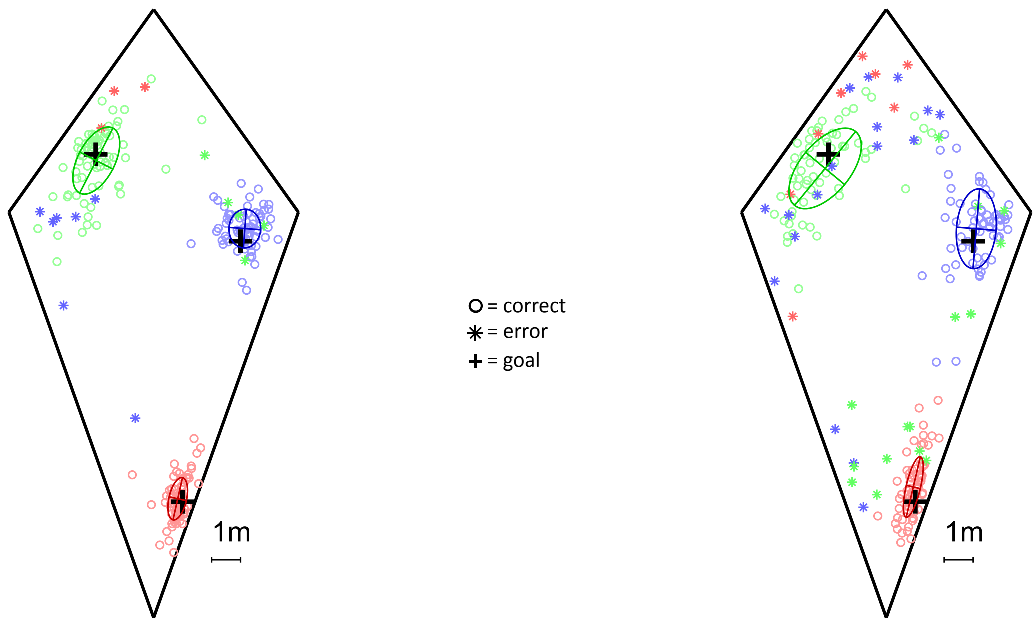
Confirmed Decision Points

First, we looked at the locations in the room where participants confirmed their decisions. Figure 2.5 shows these results comparing the texture condition to the random-dot condition for all three goals.

At a first glance, participants were able to accomplish the task in the random-dot condition, but they made more errors compared to the texture condition. Comparing the percentage of error trials between the two conditions shows that they made significantly less errors in the texture condition than in the random-dot condition (texture=6.2%, random-dots=16,67%, wilcoxon rank sum test $p<0.01$). However the error rate in the random-dot condition was still below chance level. One question might be why they still made errors in the texture condition, where they could clearly see the room. The reason for this is probably that most of these errors are rotational errors, where participants searched in the opposite location in the room.

Also in the texture condition, the decision points are more densely clustered around the respective goal position, whereas in the random-dot condition they are more widely distributed. The mean distance between the confirmed decision points and the respective correct goal position was significantly smaller in the texture condition (texture=0.8790m, random-dots=1.2945m, wilcoxon rank sum test $p<0.001$). This can be seen as well when looking at the error ellipses of the correct decisions which are larger in the random-dot condition (texture mean=1.47m^2, random-dot mean=2.78m^2).

Other than these differences between the two conditions, there are also differences between the three goal positions. In both conditions the performance is best for goal B (red goal). Here they make less errors and the distribution of the correct decisions is more dense compared to the other two goal positions. This is supposedly due to the fact that the angle of this corner is sharper than the other corners and is therefor easier to find. Participants' performance is worst for goal C (green goal) in both conditions. The reason for this could be that this goal position is not lying in or close to one of the corners but between two corners, and therefor it is harder to judge the exact position.



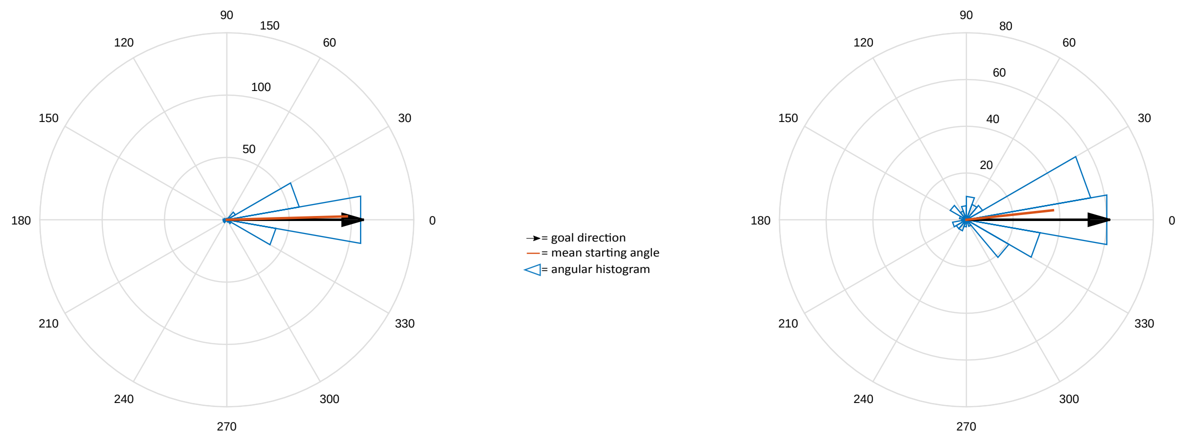
(a) texture (b) random-dot

Figure 2.5: Confirmed decision points of the baseline condition. Correct decisions are plotted as circles, errors are plotted as asterisks. Error ellipses are shown for correct decisions only. (a) Texture condition. (b) Random-dot condition.

Initial Walking Direction

Initial walking directions for the baseline experiment are shown in figure 2.6. Here the black arrow indicates the direction of the goal location. For comparison, all goal directions have been assigned 0°. The orange line represents resultant vector $\vec{r}$, i.e. the principal direction of the histogram. The length of $\vec{r}$ is a measurement for the deviation of the individual starting angles from the principal direction. The longer the resultant vector, the higher the concentration of the data around the principal direction. The variance of the starting angles is therefor given by $1 - \vec{r}$. For the texture condition there was almost no difference between the initial walking directions and the respective goal direction. A rayleigh test for non-uniformity of circular data shows that the starting angles are non-uniformly distributed ($p<0.001$). Also a v test for non-uniformity of circular data shows that the resultant vector does not differ significantly from the true goal directions ($p<0.001$). That means participants knew at least in which segment of the room the goal was and after orienting themselves in the test-phase, they walked almost always directly towards this goal segment.

Compared to the texture condition, the difference between the initial walking directions



(a) texture

(b) random-dot

Figure 2.6: Initial walking directions of the baseline experiment. Histograms are pooled over all three goals, true goal directions are set to zero (black arrow). Orange line depicts the mean resultant vector $\vec{r}$, length of $\vec{r}$ shows concentration of the distribution around the mean. (a) Texture condition. (b) Random-dot condition.

and the goal direction was a bit larger in the random-dot condition (6.26°) and the data were less concentrated around this resultant vector. Here, too, a rayleigh test did not show a uniform distribution of starting angles ($p<0.001$) and a v-test also showed no significant deviation from the respective goal direction ($p<0.001$). Participants were also able to start their search in a goal-directed way. This shows that regarding their initial idea of where the goal was and which direction to go, there was only a small difference between the texture and the random-dot condition.

Trajectories

Another characteristic of participants' search behavior that we looked at were the trajectories for each individual trial. Here trajectories are displayed for each goal with the two respective starting locations in one plot for each conditions. Figure 2.7 shows these trajectories for the baseline experiment. For reasons of clarity and comprehensibility the remaining figures of the other experiments can be found in A.2.1.

Trajectories show a clear difference in searching behavior between the texture and the random-dot condition. Although participants could solve the task quite well in the random-dot condition, their searching behavior seems to be more uncoordinated and less goal oriented than in the texture condition. In the latter condition they walked more or less straight towards the goal in most of the trials.

Again, there is also a difference between the three goal locations. For the red goal location, there seems to be a great number of trials, where they also walked straight from the start towards the goal even in the random-dot condition. That means their searching behavior for this goal location was more goal oriented than for the other two goals. The widest distribution of the trajectories can be seen for the green goal where participants clearly searched less goal oriented. This shows again that the three goal locations varied in their level of difficulty.

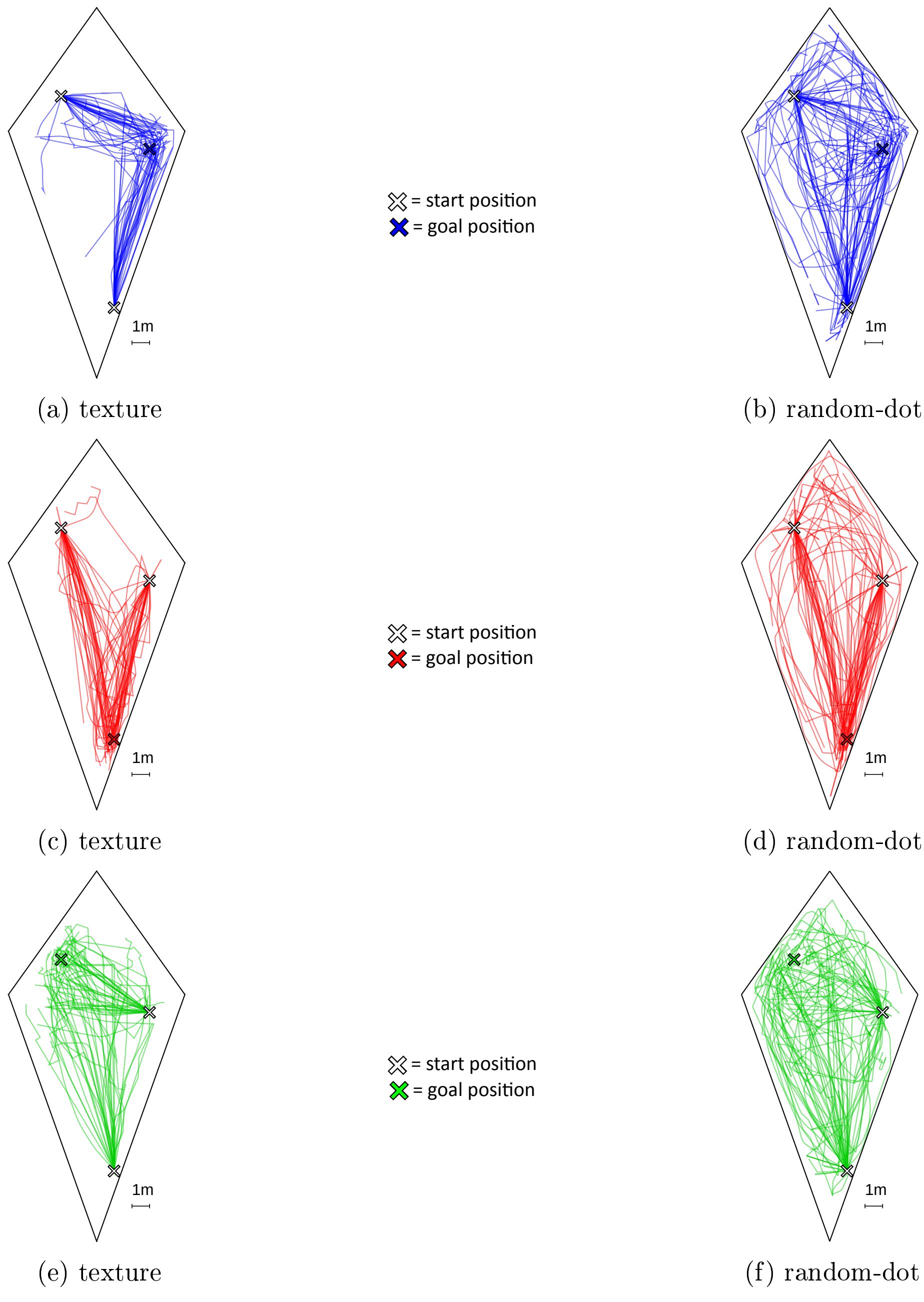


Figure 2.7: Trajectories of the baseline experiment with the stereoscope, plotted for each goal (colored cross) with the two respective start positions (white cross). (left column) Texture condition. (right column) Random-dot-condition.

Time

The time that participants spent in the test-phase searching for the goal was significantly shorter in the texture condition (texture=14.057s, random-dot=17.37s, wilcoxon rank sum test $p<0.001$). The difference between the texture and random-dot condition is even bigger for the encoding time. Here they spent significantly more time in the random-dot condition than in the texture condition (texture=13.61s, random-dot=21.07s, wilcoxon rank sum test $p<0.001$).

Preliminary Test and Correlation with Main Task

For the evaluation of the preliminary test we plotted the percentage of correct responses against the respective differences. We then fitted a psychometric curve and determined the threshold distance where they answered 75% correctly. Figure 2.8 shows this curve and the corresponding threshold (1.587m) averaged over all participants for the baseline experiment. There were three participants that had an unusual high threshold or were below chance level even at an inter-object distance of 5 meters. These outliers are not included in the results presented here. We then plotted the threshold for each individual

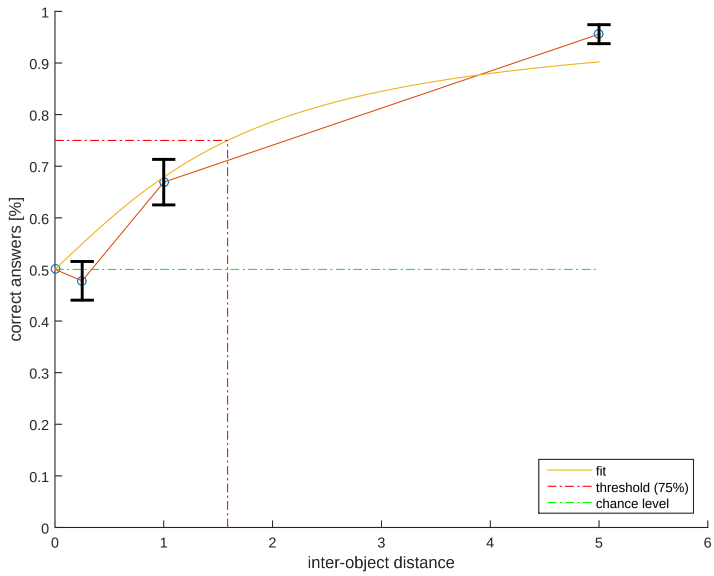


Figure 2.8: Results of the preliminary test. Percentage of errors plotted for each inter-object distance. Yellow curve shows the psychometric fit, red dotted line depicts the threshold where participants answered 75% correctly.

participant against their respective performance in the main experiment, as shown in figure 2.9. There was no correlation between performance in the preliminary test and in the random-dot task. As a consequence of this, the outliers from this preliminary test were not excluded from the analysis of the main experiment.

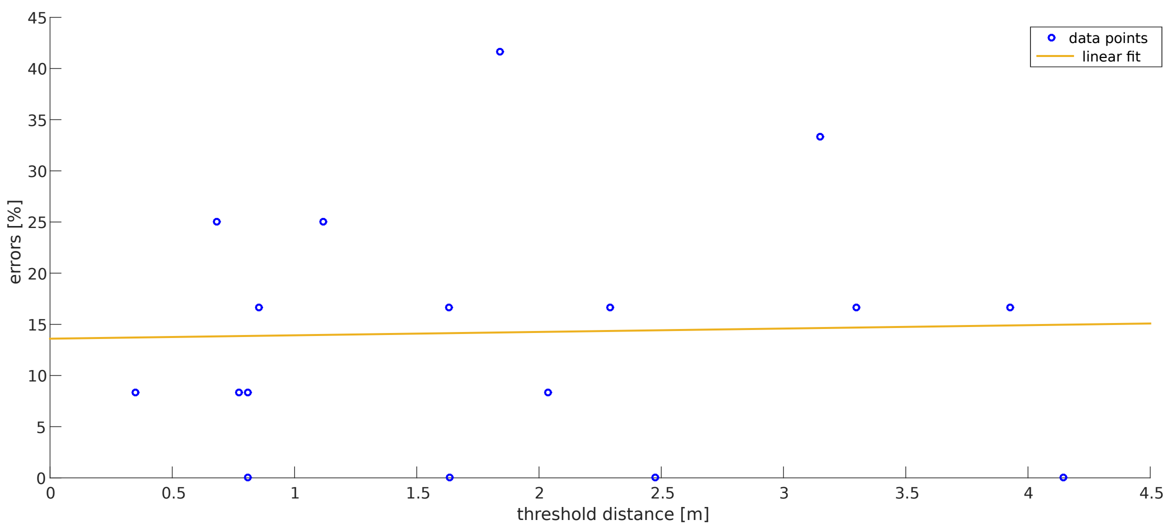


Figure 2.9: Results from the preliminary test plotted against the performance (percentage of errors) in the main experiment for each participant. Yellow line shows the linear fit.

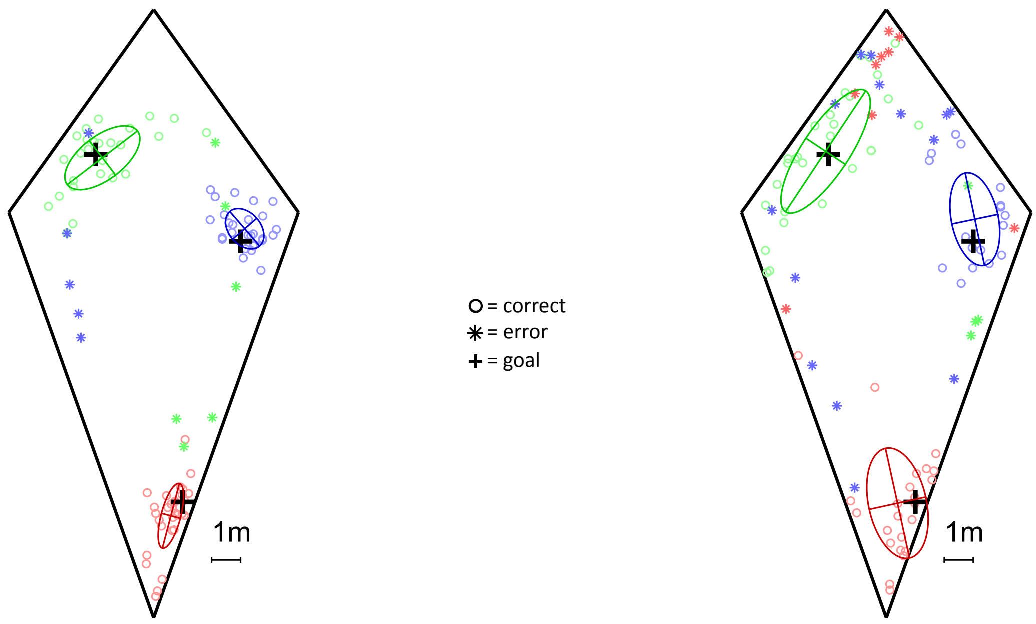
2.2.3 Experiment 2: Monocular Viewing

Confirmed Decision Points

Figure 2.10 shows the confirmed decision points in the monocular viewing experiment for both conditions. As in the baseline experiment, the error rate was higher in the random-dot condition than in the texture condition. Although the decision points are obviously wider distributed in the random-dot condition, the error rate was not significantly different (texture=28.125%, random-dot=11.46%, wilcoxon rank sum test $p>0.05$). Compared to the baseline experiment, the error rate was higher in both conditions. In both cases this difference was not significant (wilcoxon rank sum test, texture: $p>0.1$, random-dot: $p>0.1$).

Regarding the statistical error it can be said that the error ellipses are larger in the random-dot condition (texture mean=2.06m^2, random-dot mean=5.55m^2), but they are also larger when comparing the corresponding conditions in the baseline and the monocular experiment. The mean distance to the true goal location is significantly shorter in the texture condition compared to the random-dot condition (texture=1.1m, random-dot=1.92m, wilcoxon rank sum test $p<0.001$). For both conditions the mean distance is significantly longer compared to the baseline experiment (wilcoxon rank sum test, texture: $p<0.05$, random-dot: $p<0.001$).

This seems to indicate that without stereo vision participants are still more or less able to find the correct segment but they have significant problems in finding the exact location.



(a) texture (b) random-dot

Figure 2.10: Confirmed decision points of the monocular experiment. Correct decisions are plotted as circles, errors are plotted as asterisks. Error ellipses are shown for correct decisions only. (a) Texture condition. (b) Random-dot condition.

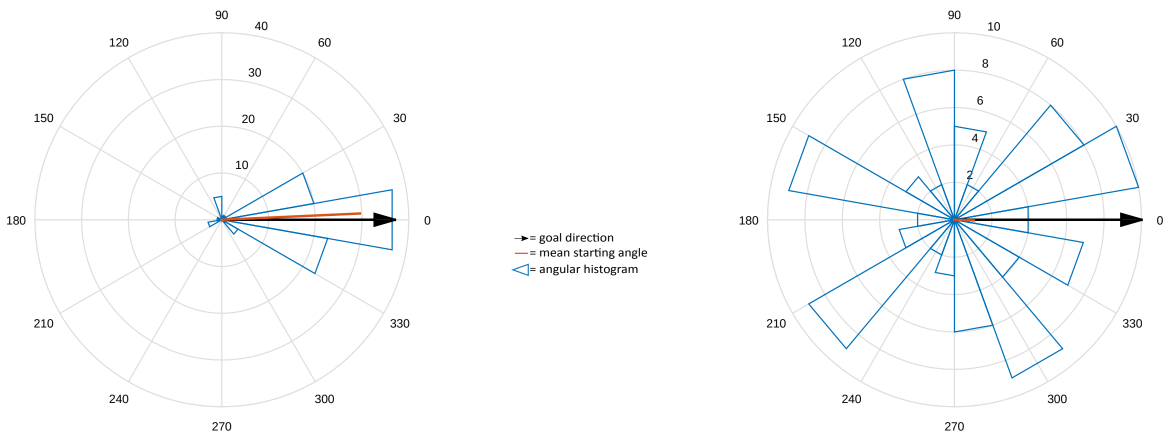
Initial Walking Direction

The initial walking directions for both conditions in this experiment are shown in figure 2.11. For the texture condition participants started searching as goal-oriented as in the baseline experiment. The initial walking directions were non-uniformly distributed (rayleigh-test $p<0.001$) and they did not differ significantly from the true goal direction(v-test $p<0.001$). The results of the random-dot condition however are completely different from the baseline experiment. Here, the initial walking directions are uniformly distributed (rayleigh-test $p>0.1$) and they also differ significantly from the true goal direction (v-test, $p>0.05$). The circular histogram for this condition clearly shows that participants seemed to have no idea where to go, as they started walking in almost every direction.

These results suggest that with motion parallax alone it is not only harder to find the goal location but also the initial orientation within the room is significantly impaired.

Trajectories

The trajectories for this experiment are shown in the appendix (A.2.1). For the texture condition the trajectories are comparable to those from the baseline experiment, showing a clearly goal-directed searching behavior, especially for the red goal. In contrast to this, the trajectories are running through the whole room in the random-dot condition. This again indicates that participants somehow managed to approach the correct goal segment,



(a) texture (b) random-dot

Figure 2.11: Initial walking directions of the monocular experiment. Histograms are pooled over all three goals, true goal directions are set to zero (black arrow). Orange line depicts the mean resultant vector $\vec{r}$, length of $\vec{r}$ shows concentration of the distribution around the mean. (a) Texture condition. (b) Random-dot condition.

but they seemed to have no clear strategy in finding the goal.

Time

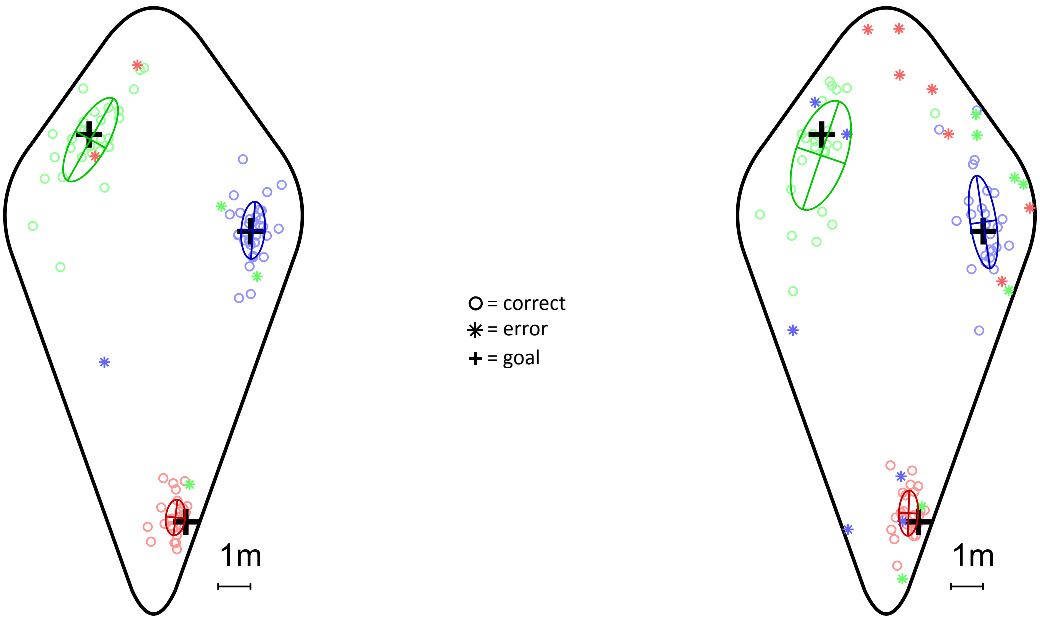
As in the baseline experiment, the search times were significantly shorter for the texture condition than for the random-dot condition (texture=16.356s, random-dot=20.504s, wilcoxon rank sum test $p<0.05$). Compared to the baseline experiment the search times were longer for both conditions without being significant (wilcoxon rank sum test, texture: $p>0.05$, random-dot: $p>0.05$). The encoding time was significantly longer in the random-dot condition (texture=19.369s, random-dot=30.513s, wilcoxon rank sum test $p<0.001$). Also the encoding times were significantly longer than in the baseline experiment (wilcoxon rank sum test, texture: $p<0.001$, random-dot: $p<0.001$).

2.2.4 Experiment 3: Rounded Corners

Confirmed Decision Points

The confirmed decision points for the rounded corners experiment are shown in figure 2.12. Again the error rate was higher in the random-dot condition compared to the texture condition (texture=6.25%, random-dot=$20.8\overline{3}$%, wilcoxon rank sum test $p>0.5$). Compared to the results of the baseline experiment the error rate was higher in the random-dot condition without being significant (wilcoxon rank sum test $p>0.1$). In the texture condition the mean error rate was the same in both experiments (wilcoxon rank sum test p>0.5).

For the statistical error, the mean distance to the goal location was longer in the random-dot condition, but this difference was not significant (texture=0.843m, random-dot=1.137m, wilcoxon rank sum test p>0.05). Interestingly, when compared to the baseline experiment the mean distance was shorter for both conditions, while being only significant in the random-dot condition (wilcoxon rank sum test, texture: p>0.1, random-dot: p<0.05).

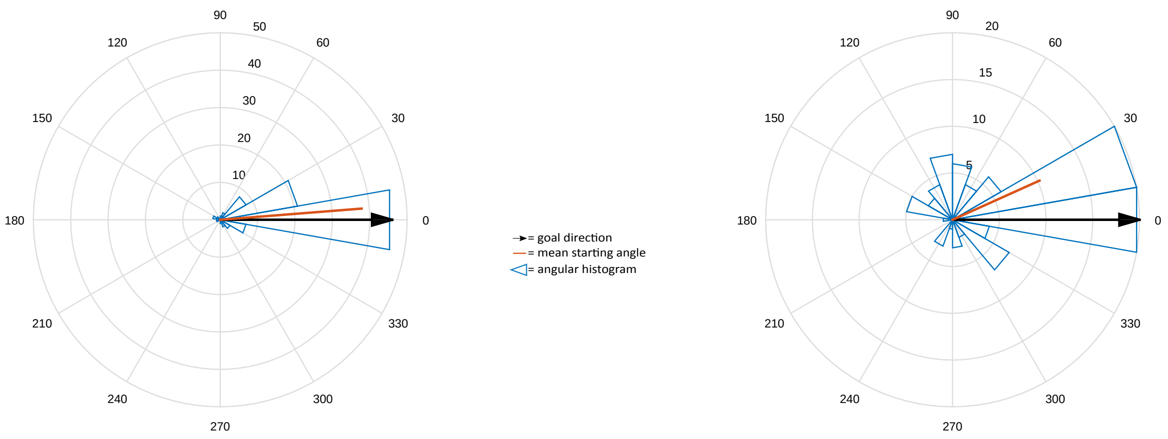


(a) texture (b) random-dot

Figure 2.12: Confirmed decision points of the rounded corners experiment. Correct decisions are plotted as circles, errors are plotted as asterisks. Error ellipses are shown for correct decisions only. (a) Texture condition (b) Random-dot condition.

Initial Walking Direction

Figure 2.13 shows the circular histograms of the initial walking directions for this experiment. In general the results are comparable to the baseline experiment. For both conditions the initial walking directions are distributed non-uniformly (rayleigh-test, texture: p<0.001, random-dot: p<0.001). Also a v-test shows no significant deviation from the true goal direction for both conditions (texture: p<0.001, random-dot: p<0.001). This indicates that participants started searching the same way as they did in the baseline experiment.



(a) texture (b) random-dot

Figure 2.13: Initial walking directions of the rounded corners experiment. Histograms are pooled over all three goals, true goal directions are set to zero (black arrow). Orange line depicts the mean resultant vector $\vec{r}$, length of $\vec{r}$ shows concentration of the distribution around the mean. (a) Texture condition. (b) Random-dot condition.

Trajectories

The trajectories for this experiment are shown in figures A.7-A.9 in A.2.1. The results here are comparable to the trajectories from the baseline experiment. For the texture condition participants' searching behavior is quite goal-directed. Especially in trials with the red goal they walk straight to the target location, whereas for the green goal the trajectories are a bit more scattered. In the random-dot condition participants seem to search less goal-directed compared to the texture condition. Here however the difference between the three goals is not that striking.

Time

As in the previous experiments participants needed more time to find the goal in the random-dot condition but this difference was not significant (texture=19.139s, random-dot=23.231s, wilcoxon rank sum test $p>0.1$). Also, the search times were significantly longer when compared to the baseline experiment (wilcoxon rank sum test, texture: $p<0.001$, random-dot: $p<0.01$). The encoding times were shorter in the texture condition but not significantly (texture=18.38s, random-dot=23.024s, wilcoxon rank sum test $p>0.1$). When compared to the baseline experiment the encoding times were shorter for both conditions. This difference was only significant in the texture condition (wilcoxon rank sum test, texture: $p<0.001$, random-dot: $p>0.5$).

2.3 Discussion

Baseline

The question was, if place recognition is possible, when the layout of the environment can only be inferred from depth perception. The results of the baseline experiment indicate that this is indeed possible.

Analyzing the confirmed decision points shows that participants made significantly more qualitative errors in the random-dot condition. Also the statistical error was significantly larger, as can be seen from the error ellipses as well as from the mean distances between the decision points and the true goal location. These findings were as expected, regarding the fact that the texture condition offered a much broader variety of cues, which participants could make use of. However, the error rate is still distinctly below chance level and the qualitative errors can partially be explained by the participants confusing the geometrically equivalent opposite corner, resulting in rotational errors. This means that in these cases they memorized geometric properties of the goal location, but the pure depth information did nor always provide enough cues to distinguish between the two alternatives. It is also to be noted that estimating distances in virtual environments is difficult and that people seem to underestimate distances (Thompson et al., 2004), even when using a stereoscopic HMD system (Loomis and Knapp, 2003). This might be one explanation for the higher statistical error in the random-dot condition.

Participant's searching behavior was analyzed by looking at the initial walking directions and the trajectories. In both conditions the initial walking direction over all participants did not differ significantly from the true goal direction. This shows that even in the random-dot condition participants had an idea where at least the correct goal segment was located. After orienting themselves they started walking more or less straight towards that segment. The trajectories show a difference between the two conditions. Here in the texture condition the searching was more goal-directed whereas in the random-dot condition participants' paths are wider distributed. This indicates that finding the correct segment was well manageable but finding the exact goal position was more difficult.

The time participants needed for each trial was significantly longer in the random dot condition, for the encoding time as well as for the search time. This shows that in the random-dot condition a longer time is needed to build up a spatial representation. Also the longer search time might indicate a higher level of uncertainty about the exact goal location.

One interesting finding that is to be mentioned, is the difference in difficulty between the three goal positions. For all variables that were analyzed, the results indicate that the red goal location was easier to memorize than the other two goals. Here participants made less errors and were closer to the correct target location but also their searching behavior was more goal-directed as can be clearly seen from the trajectories. This holds for both conditions, but it is especially noticeable that the difference between the random-dot and the texture condition was smallest for the red goal. Except for a few rotational errors, participants' seemed to have no problem in finding this goal based on depth cues only. The obvious reason for this is that the red goal lies very close to the sharpest corner of the room, a location that is easy to remember and to distinguish from other locations. The most difficult goal seemed to be the green one. This goal is not located close to a specific corner but rather in the mid-distance between two corners, making it harder to estimate the exact position.

Finally we compared the results of the preliminary test with the performance in the main experiment. Results from the pre-test showed that participants could perceive stereo vision and were able to distinguish the two squares. However there was no correlation between stereo ability and the performance in the main experiment. There are two possible explanations for this finding. First, participants that had a lower stereo ability might rely on motion parallax in the random-dot condition to solve the task, allowing them to achieve comparable performance. The influence of motion parallax was tested in the monocular experiment. Second, in the pre-test participants only needed to judge relative distances, whereas in the main experiment the task demanded an absolute estimation of distances. Therefor, a higher stereo ability measured in the pre-test must not necessarily lead to a higher performance in the random-dot condition.

Taken together, the results presented here indicate that place recognition is possible, solely based on depth information but that performance increases as more cues become available. The question if disparity or motion parallax is the stronger cue is not of importance in this context, as both are taken together as depth cues here.

Monocular

Given the results of the baseline experiment, the question was how important is motion parallax for place recognition, or can a location be memorized even without disparity? In the second experiment described here, using a monocular viewing condition, participants again made more errors in the random-dot compared to the texture condition as expected. When compared to the baseline experiment, the percentage of qualitative errors was higher for both conditions. Also the statistical error was larger, with the mean distance to the true goal position being significantly longer in both conditions. Especially for the random-dot condition this difference was highly significant, and it can be seen that the decision points are much broader distributed. This shows that participants somehow managed to find the correct segment, but the exact goal location was hard to find.

The searching behavior for the texture condition did not differ so much from the baseline experiment. Initial walking directions as well as the trajectories show a clear goal-directed searching. However, in the random-dot condition the initial walking directions are uniformly distributed around the whole circle and the trajectories are covering almost the whole room. This indicates that participants had no initial idea of where the correct segment was located.

Furthermore the encoding time as well as the search time were longer compared to the baseline experiment. This difference was significant for the encoding time, indicating that it was a lot more difficult for participants to orient themselves monocularly in the random-dot room. The fact that the difference for the search time was not significant might be due to the low number of participants (n=8). Another plausible explanation might be motivation. Doing this task monocularly was very difficult and stressful, causing participants to shorten the trial by confirming to early.

It is to be mentioned that motion parallax is of course a strong cue for depth perception. One can clearly recognize objects and judge relative distances from random-dot patterns when performing sidesteps in virtual reality. For completing the homing task in the present experiment however, motion parallax is only minimal helpful. Participants could estimate their position relative to the goal by moving left- or rightward. But during forward movement or, even worse, during rotations, they were completely lost. Thus they had to alternate between estimating the goal location and moving towards this assumed target. This made the task in the monocular experiment significantly more difficult. In other words, building up a spatial image using motion parallax is possible, but for spatial updating of this representation during movement disparity is needed.

Altogether, the results of the monocular experiment showed that it is at least possible to find the vicinity of a learned location but that this ability is greatly impaired without disparity. It takes longer to encode a location and the searching is more prone to failure. This indicates that stereo vision is indeed a strong cue for place encoding and spatial updating.

Rounded Corners

The first experiment showed that navigation and place recognition is possible based on depth information only. Although we tried to eliminate all other cues, one could still argue that the sharp corners in the kite-shaped room might serve as some kind of landmark information.

The results of the experiment with the rounded corners showed that the error rate was comparable to the baseline experiment. No significant difference was found for both conditions. The statistical error was even smaller with rounded corners, being significant in the random-dot condition. This shows that despite not being able to see the corners anymore, participants could still accomplish the task without problems.

The initial walking direction and the trajectories were comparable to the baseline experiment as well, showing a clearly goal-directed searching behavior. This shows that the rounded corners had no effect on participants ability to orient themselves within the room.

Regarding the times spent, the search time was longer compared to the baseline experiment. This might be due to the fact that the rounded corners induce a higher level of uncertainty about the estimated position. In contrast, the encoding times were shorter than in the baseline experiment. This contradiction between the search and encoding times might be due to the fact that the rounded corners can still easily be seen from a more distant point of view, by looking at the edge of the ceiling. This allowed participants to easily orient themselves and get an idea of were the goal was located. However when getting closer to a wall it became harder to see the ”corner”, making the exact goal location more difficult to find.

All these results of the third experiment show that eliminating potential landmark information by rounding the corners did not impair participants navigational abilities. It can therefore be concluded that place recognition is possible based on depth cues only without the need for landmark information.

Chapter 3

Place Recognition in an Immersive VR Setup

This chapter deals with another two experiments that were carried out to investigate the question asked in chapter 2. The difference is that whereas in the first series of experiment participants were seated in front of the stereoscope to view the virtual environment, this time they were wearing a head-mounted display (HMD).

We repeated the baseline experiment and the experiment with the rounded corners as described above. The virtual environment itself and the procedure of the experiments were as explained in 2.1.2.

3.1 Reasons for Using a Head Mounted Display

The question is, given the advantages of the stereoscope as stated in 1.3.2, what are the benefits of repeating the experiments with an HMD?

The first reason is that wearing an HMD gives the participant a wider field of view both horizontally and vertically. This results in a higher immersion and a stronger presence as compared to viewing the environment on the stereoscope. The monitors used in the setup of our stereoscope were running at a much higher resolution than the display built in the HMD (2520x1280 pixels vs. 960x1080 pixels per eye). However, the impression of actually "being" inside the scene was stronger on the HMD due to the larger field of view. It is also to be noted that for the random-dot pattern it is not necessarily required to have a high resolution, as the dot size is set to few pixels only.

Second, the participants were able to look around freely by turning their head when wearing the HMD, providing additional proprioceptive feedback. On the one hand, this enhanced the presence, making navigation within the environment feeling more natural.

On the other hand, the addition of proprioceptive feedback is expected to help them significantly to orient themselves within the virtual environment.

A third reason arises from the following observation. The same random-dot environment was shown to some colleagues, using the same parameters (dot lifetime, number of dots etc.) as in the stereoscope setup. Surprisingly they reported that they were almost immediately able to perceive the room. In contrast to this, on the stereoscope there was always a short moment that participants needed to fuse the two images and perceive a 3D impression.

Finally the shorter distance between the observer and the monitor on the HMD as compared to the viewing distance on the stereoscope brought participants "closer" to the scene. This might allow for a better estimation of distances to the walls and thus a better judgment of ones current position based on depth information.

Taken together these factors are expected to increase presence and thus participants' impression of actually being in the room. The ability to orient within this virtual room should increase as well. Therefor participants' are expected to have a better knowledge about their position relative to the goal location, improving their navigational performance in the homing task. Also the time needed to encode the current goal location is expected to decrease.

3.2 General Methods

3.2.1 Setup

Apparatus

For the reasons described in 3.1 we used a head-mounted display in the following two experiments. The Device used here was an Oculus™ Rift DK2 from Oculus™ VR. DK stands for development kit, meaning that this is not a consumer product, it is currently under development and available for software developers only. The "2" indicates that this is the second development kit that is available and which has been significantly improved compared to the first version.

The second version uses a higher resolution by integrating a full hd display (1920x1080 pixels) for both eyes, resulting in a resolution of 960x1080 pixels per eye. In contrast the first version only had 640x800 pixels per eye, so visual quality is greatly improved. Another problem with the original version is concerned with latency and tearing. That means that when the user performed fast head movements the image became blurred. This could easily induce nausea. The DK2 uses an OLED (organic light emitting diode) display with a very low reaction time (≪1ms). Together with a so-called "low persistence" technology this tearing effect is significantly decreased. Lastly the second version has a reduced number of cables and ports that needed to be connected to the computer. Whereas the original version had an external box that required a separate power supply, the DK2 is just connected via thin Hdmi and USB cables. This greatly increased the wearing comfort.

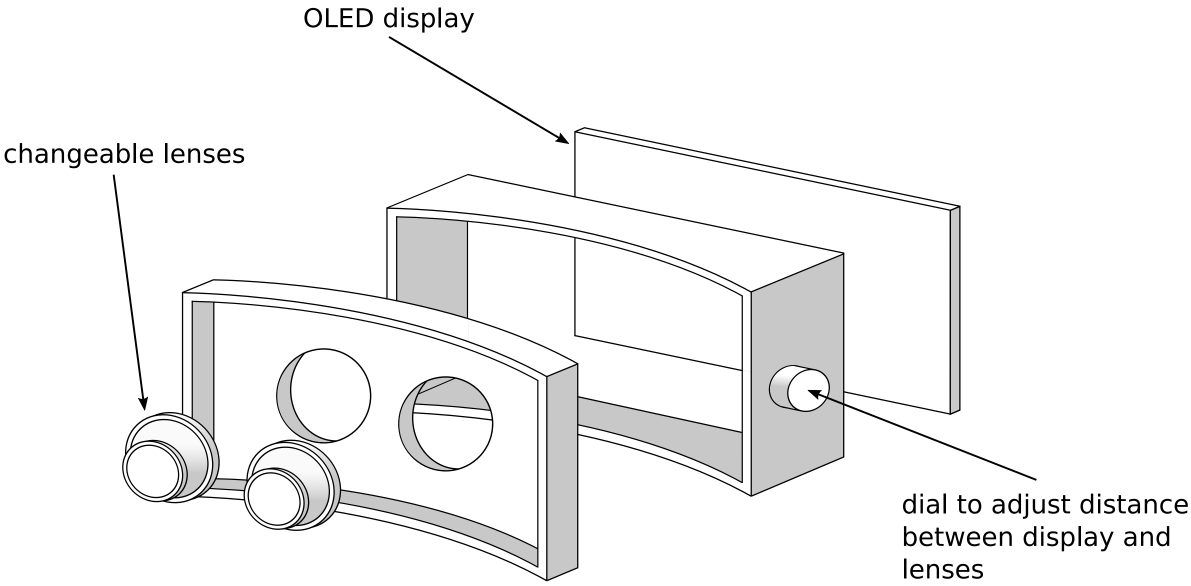


Figure 3.1: Schematic of the Oculus Rift DK2 redrawn from Parkin (Parkin and Macneill, 2014).

Figure 3.1 shows a schematic of the Oculus Rift. The built-in display is a 5.7 inch screen with a resolution of 1920x1080 pixels running at a refresh rate of 75 Hz. As the distance between the screen and the eyes is rather short (4.1 cm), there are two fixed-focus lenses to show a sharp image to the user. The short viewing distance in combination with the lenses results in a wide field of view (fov) of about 84°(horizontal). Because of the characteristics of the Oculus Rift, the software has to be adjusted to present a proper image and to achieve the desired 3D impression.

The most obvious adjustment is that the scene has to be "doubled". Therefore the screen is virtually split up into two halves, one for each eye. Each half then displays an image of the scene which is shifted by a small amount to account for the difference in eye position. Two other adjustments that have to be made are due to the optical properties of the lenses. First, the image seen by the user is of course distorted. This so-called pincushion distortion has to be corrected by applying a barrel distortion to the rendered scene. Second, the lenses cause a chromatic aberration, i.e. the lenses do not focus all colors to the same convergence point. This effect has to be corrected by the software by shifting the colors in the image accordingly. Figure 3.2 shows these corrections for a simple demo scene.

The Oculus Rift has a built-in 3-axis gyroscope, accelerometers and magnetometers to track the head orientation. Data from these sensors is tracked with 1000 Hz and sent to the PC via USB cable so that the camera position in virtual reality is updated accordingly with almost no latency. To measure head position for positional tracking, the second version of the Oculus Rift has infrared LEDs built-in. These LEDs are tracked by an infrared camera that can be placed in front of the user. The disadvantage of this current configuration of LEDs is that they are only placed on the front of the HMD and the camera can only "see" what is within its fov (60°). This means that whenever a user turns his head too far, the LEDs move out of the camera's fov, and there is no more tracking possible. The consequence is that as the user turns his head too far, suddenly the simulation no longer reacts to head translations. This interruption feels very unnatural and irritating to the user, even more unnatural than not having positional tracking at all. Therefore we decided not to use positional tracking. Thus participants did not have the possibility to use motion parallax from head movements, but they still could make sidesteps to the left or to the right to get information from motion parallax. The experiment was carried out on an Intel® Core™ i5-3470 @ 3.2 GHz with 8 GB working memory and an Nvidia® Geforce® GTX™ 560 Ti GPU.

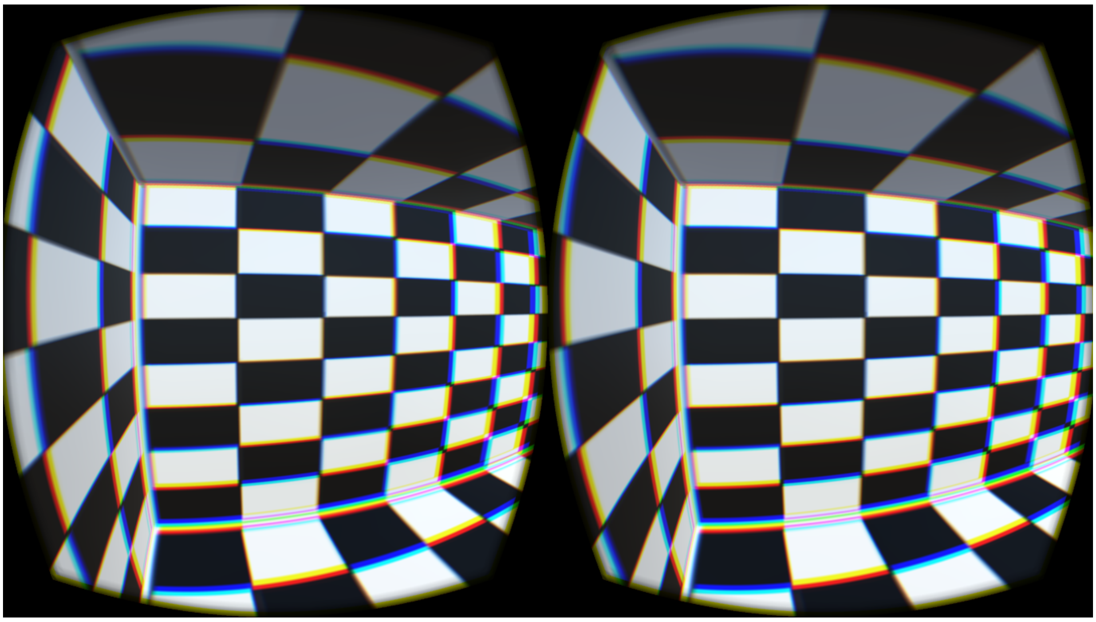


Figure 3.2: Example scene showing the barrel distortion and the correction for the chromatic aberration.

Virtual Environment

The environment used was the same kite-shaped room as in the previous experiments. For the texture condition also the same texture was placed on the walls. In the random-dot condition the parameters of the random-dot renderer were adjusted to account for the display differences of the HMD. First, the size of the dots was decreased from 3 pixels to 1 pixel. The decrease in size was due to the fact that the lenses built in the HMD led to a distortion of the dots that made them look like drops in the peripheral field of view. This effect was reduced by decreasing the size of the dots. Also the maximum number of dots was increased to 3000. This was done to account for the wider fov, keeping the density of the random-dot pattern comparable to the stereoscope experiments.

Participants should be given the advantage of being able to turn and look around freely within the virtual environment during a trial with the HMD. Therefor they were no longer sitting on a chair but rather standing during the whole experiment. To move around they used a wireless controller (Xbox® 360 wireless controller for Windows®). They could control the forward and backward or left and right (sidesteps) movement with the left analog stick of the controller. The direction of movement was controlled by participants' head direction, i.e. they were always moving forward in the direction that they were looking. This way of navigating in the virtual environment gave participants additional proprioceptive feedback but also created a stronger presence as described above.

Two different experiments were conducted using this setup. In the first experiment the standard room as in the stereoscope baseline experiment was used. In the second experiment the corners of the room were rounded as in 2.1.3. There was no monocular setup this time.

Software

The software used in the HMD experiments was basically the same as in the previous experiments, there was no major change to the program sequence. However, to display the environment on the Oculus Rift, a few changes to the software had to be made.

First, the Oculus Rift had to be integrated into the existing software project. This was done by using the software development kit (Oculus™ SDK for Windows V0.4.4-beta) provided by Oculus™ VR. This SDK provides all the necessary interfaces and libraries to implement virtual reality applications, for example, it allows to read out the current state of the sensors to get the users' head rotation.

Second, the OSG viewer had to be adjusted to account for the optical properties of the Oculus Rift as shown in figure 3.2. To achieve this the *osgoculusviewer* written by Björn Blissing and Jan Ciger (https://github.com/bjornblissing/osgoculusviewer.git) was used. This software was published under BSD license and consists of a set of classes containing all interfaces and features needed to integrate the Oculus Rift into existing projects. The *osgoculusviewer* was then adjusted to run with the random-dot pattern generator used here.

A third change concerned the way participants controlled their movement within the virtual environment. In the previous studies, participants controlled forward and sidewards movement with the keyboard and could turn around by moving the mouse. This was implemented by having a virtual camera and then apply the input from mouse and keyboard to this camera. As mentioned above the idea for the HMD setup was to let participants look around freely by turning their head and move forward in the direction that they were looking at. Thus the rotation and position of the virtual camera had to be updated every frame according to the current heading direction. The camera rotation itself was updated directly by the *osgoculusviewer*. To update the position, the rotation of the HMD was retrieved as a quaternion, here an *osg::Quat* q=(q_x, q_y, q_z, q_w). From this quaternion the current heading direction was calculated as $heading$=$atan2\big(2.0 * (q_x * q_y + q_z * q_w), (q_x^2 - q_y^2 - q_z^2 + q_w^2)\big) * (\frac{180}{\pi})$. The heading direction (in degrees) together with the step range (forward translation relative to the joystick input) was used to calculate the offset by which the camera position was updated was as $offset_x = step * sin(heading * \frac{\pi}{180})$ and $offset_y = step * cos(heading * \frac{\pi}{180})$ respectively.

3.2.2 Procedure

In general the procedure was the same as in the experiments with the stereoscope. Participants were informed about the procedure of the experiment, the data collected and the potential danger of motion sickness due to wearing the HMD. The experimenter then placed the HMD on their heads and they were placed in the middle of the lab to ensure that they had enough free space to turn. Also the experimenter had to take care during the whole time that participants did not get entangled in the cable leading from the HMD to the PC.

Participants then started the first trial by pressing a button on the controller. As with the stereoscope, they began with the texture condition and did the preliminary test followed by the random-dot condition on a different day. At the end of each day they were given the same questionnaire as in the previous experiments.

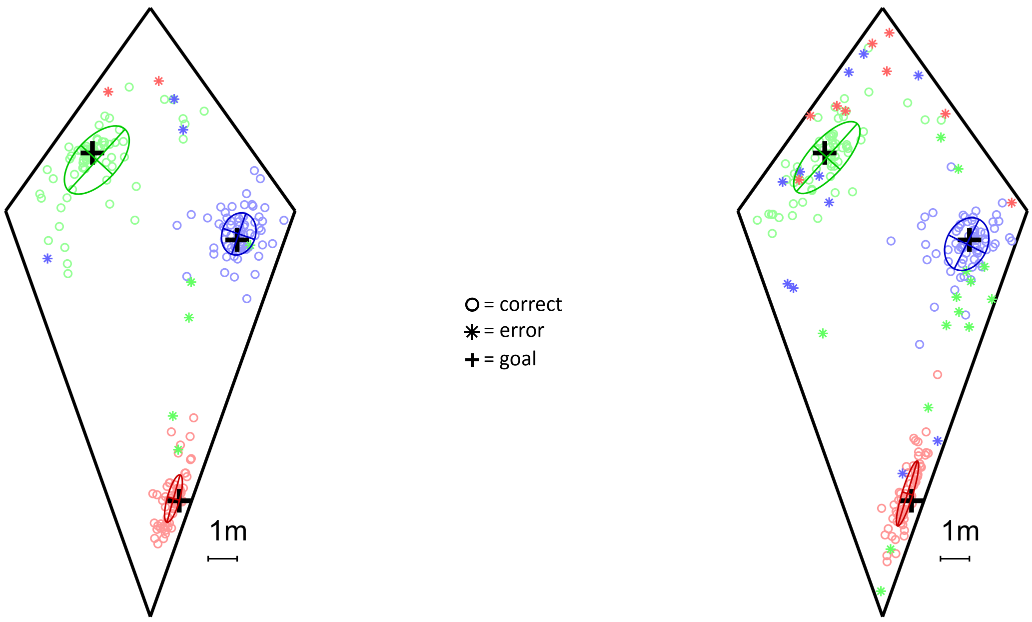
3.3 Results

3.3.1 Experiment 4: HMD Baseline

Confirmed Decision Points

Figure 3.3 shows the confirmed decision points for the HMD baseline experiment. As in the experiment with the stereoscope participants made more errors in the random-dot condition, but this difference was not significant (texture=$4.1\overline{6}$%, random-dot=$14.1\overline{6}$%, wilcoxon rank sum test p>0.1). When comparing the results of the two experiments it can be seen that in the HMD experiment they made less qualitative errors, but this difference was not significant(wilcoxon rank sum test, texture: p>0.5, random-dot: p>0.1).

Regarding the statistical error the error ellipses are slightly smaller in the texture condition (texture=1.77m^2, random-dot=1.96m^2). Also the distance to the true goal location is significantly greater in the random-dot condition (texture=0.856m, random-dot=1.086m, wilcoxon rank sum test p<0.001). For both conditions these distances are significantly smaller compared to the stereoscope experiment (wilcoxon rank sum test, texture: p<0.05, random-dot: p<0.01). This indicates that participants' performance was indeed better with the HMD where they could estimate the goal location more precisely.

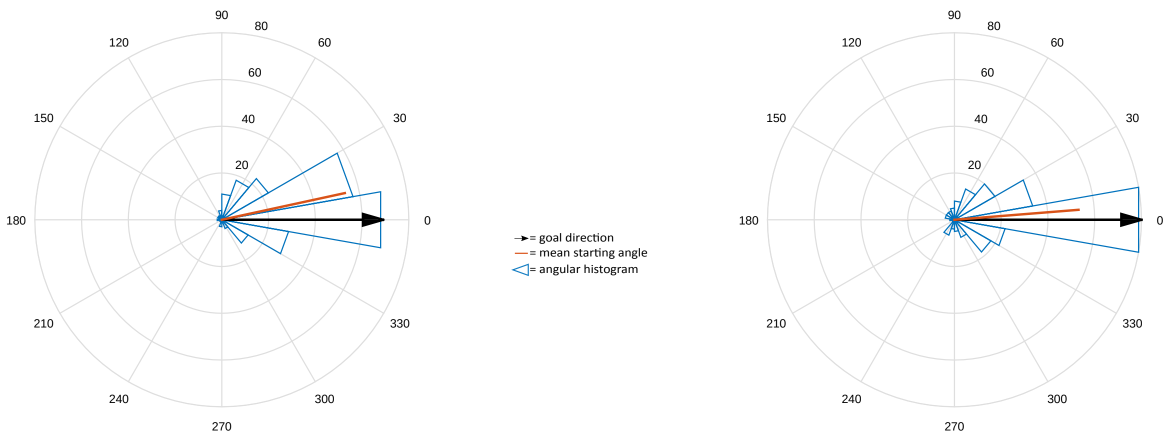


(a) texture (b) random-dot

Figure 3.3: Confirmed decision points of the HMD baseline experiment. Correct decisions are plotted as circles, errors are plotted as asterisks. Error ellipses are shown for correct decisions only. (a) Texture condition. (b) Random-dot condition.

Initial Walking Direction

Participants searching behavior is again analyzed first by looking at the initial walking directions shown in figure 3.4. For both conditions a rayleigh test shows a non-uniform distribution (texture: p<0.001, random-dot: p<0.001). Also for both conditions the initial walking directions are clearly goal oriented with a slightly wider distribution around the true goal direction in the texture condition. A v-test shows no significant difference between the starting angles and the true goal direction (texture: p<0.001, random-dot: p<0.001).



(a) texture

(b) random-dot

Figure 3.4: Initial walking directions of the HMD baseline experiment. Histograms are pooled over all three goals, true goal directions are set to zero (black arrow). Orange line depicts the mean resultant vector $\vec{r}$, length of $\vec{r}$ shows concentration of the distribution around the mean. (a) Texture condition. (b) Random-dot condition.

Time

Participants spent more time searching for the goal in the random-dot condition compared to the texture condition but this difference was not significant (texture=14.147s, random-dot=15.827s, wilcoxon rank sum test $p>0.05$). Compared to the stereoscope experiment the HMD search time was shorter in the random-dot condition, but slightly longer in the texture condition. This difference was significant only for the random-dot condition (wilcoxon rank sum test, texture: $p>0.1$, random-dot: $p<0.001$).

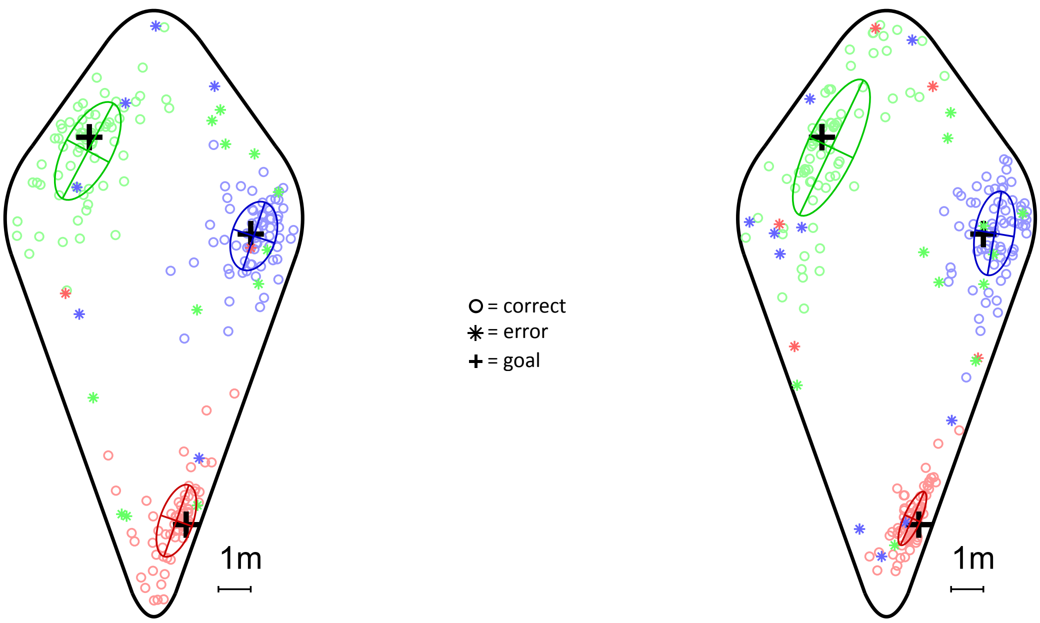
The encoding time was shorter in the texture condition without being significant (texture=12.72s, random-dot=14.7s, wilcoxon rank sum test $p>0.5$). What is interesting is the comparison of encoding times between the stereoscope and the HMD experiment. For the texture condition there is no significant difference as expected (wilcoxon rank sum test $p>0.1$). For the random-dot condition however, the encoding times were significantly shorter (wilcoxon rank sum test $p<0.001$) in the HMD experiment.

3.3.2 Experiment 5: HMD Rounded Corners

Confirmed Decision Points

Figure 3.5 shows the confirmed decision points for the HMD experiment with rounded corners. Again participants made more qualitative errors in the random-dot condition but here the difference is not significant (texture=$8.\overline{3}\%$, random-dot=$10.8\overline{3}\%$, wilcoxon rank sum test $p>0.5$). Compared to the HMD baseline experiment the error rate was higher in the texture but lower in the random-dot condition (wilcoxon rank sum test, texture: p>0.1, random-dot: p>0.5). Also when compared to the stereoscope experiment with rounded corners for both conditions the error rate was smaller for the HMD but the differences are not significant (wilcoxon rank sum test, texture: p>0.5, random-dot: p>0.5).

The analysis of the statistical error shows that the mean distance to the true goal location is significantly shorter in the texture condition (texture=1.146m, random-dot=1.345m, wilcoxon rank sum test p<0.05). For the texture as well as for the random-dot condition the mean distances are shorter compared to the stereoscope experiment. This difference was onlytrim significant for the texture condition (wilcoxon rank sum test, texture: p<0.05, random-dot: p>0.05. For both conditions the mean distances are significantly longer compared to the HMD baseline experiment (wilcoxon rank sum test, texture: p<0.001, random-dot: p<0.05).



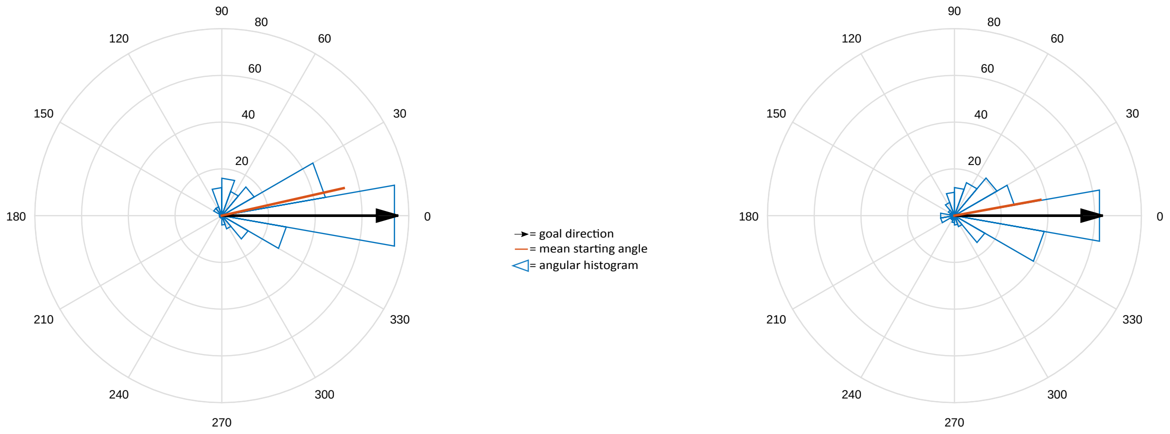
(a) texture

(b) random-dot

Figure 3.5: Confirmed decision points of the HMD experiment with rounded corners. Correct decisions are plotted as circles, errors are plotted as asterisks. Error ellipses are shown for correct decisions only. (a) Texture condition. (b) Random-dot condition.

Initial Walking Direction

The initial walking directions for this experiment are again shown as circular histograms in figure 3.6. As for the baseline HMD experiment the initial walking directions are non-uniformly distributed in both conditions (rayleigh test texture: p<0.001, random-dot: p<0.001). Also for both conditions the starting angles do not differ significantly from the respective true goal direction(texture: p<0.001, random-dot: p<0.001). That means that in this experiment again the initial walking directions are clearly goal-directed in the texture and the random-dot condition.



(a) texture (b) random-dot

Figure 3.6: Initial walking directions of the HMD experiment with rounded corners. Histograms are pooled over all three goals, true goal directions are set to zero (black arrow). Orange line depicts the mean resultant vector $\vec{r}$, length of $\vec{r}$ shows concentration of the distribution around the mean. (a) Texture condition. (b) Random-dot condition.

Time

The mean search time was shorter in the random-dot condition without being significant (texture=15.06s, random-dot=12.317s, wilcoxon rank sum test p>0.1). For both conditions when using the HMD the search times where significantly decreased in comparison to using the stereoscope (wilcoxon rank sum test, texture: p<0.001, random-dot: p<0.001). Compared to the HMD baseline condition the search time was longer in the texture condition but shorter in the random-dot condition (wilcoxon rank sum test, texture: p>0.5, random-dot: p<0.01).

Also the encoding time in the random-dot condition was significantly shorter than in the texture condition (texture=15.172s, random-dot=11.924s, wilcoxon rank sum test p<0.01). Comparing the encoding times between HMD and stereoscope experiment, in both conditions the times are significantly shorter for the HMD (wilcoxon rank sum test, texture: p<0.001, random-dot: test p<0.001). In comparison to the HMD baseline experiment, the encoding times were longer in the texture condition but shorter in the random-dot condition (wilcoxon rank sum test, texture: p>0.5, random-dot: p<0.05).

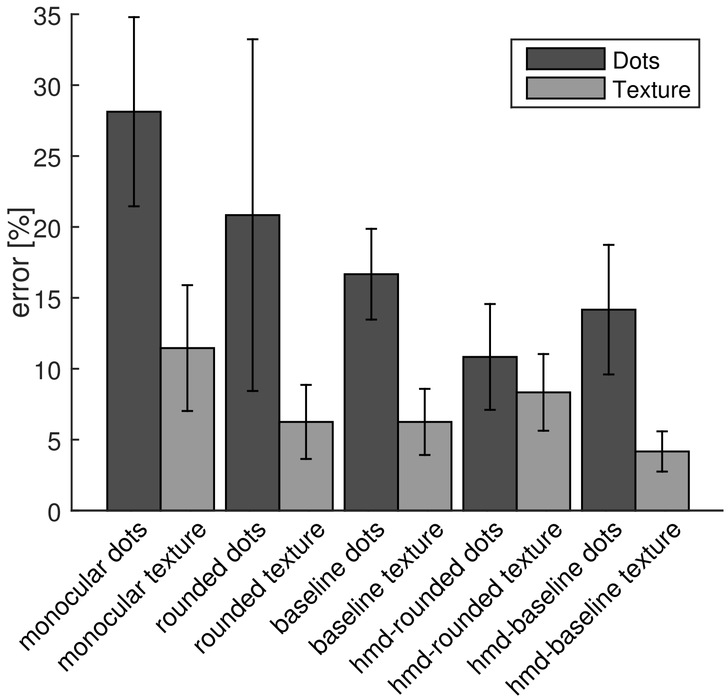


Figure 3.7: Comparison of percentage of errors for all experiments. From left to right: stereoscope monocular, stereoscope rounded corners, stereoscope baseline, HMD rounded corners, HMD baseline. Light gray=texture condition, dark gray=random-dot condition.

3.3.3 Comparison of all Experiments

In this section the overall results of the stereoscope are compared to the results of the HMD. Figure 3.7 shows a comparison of the percentage of qualitative errors for all five experiments. For the texture condition the performance is comparable over all experiments with the highest error rate in the monocular condition and the best performance in the HMD baseline condition. When looking at the random-dot condition there is a decrease in error rate from the highest error rate being in the monocular condition (28.125%) to the lowest in the HMD rounded corner condition ($10.8\overline{3}$). Therefor it can be said that the performance was indeed better when using the HMD instead of the stereoscope as we would have expected. However, it seems to be surprising that the participants made less errors with rounded corners as compared to the room with sharp corners. Also the differences in performance between HMD and stereoscope experiments were small and not significant.

A comparison for the statistical error over all five experiments can be seen in figure 3.8 where the average distance to the goal is plotted for each experiment. Here the performance seems to be comparable over all experiments for both conditions except for the random-dot condition in the monocular experiment where the statistical error was distinctly higher than in all other cases. Also in general the difference between the texture and the random-dot condition was higher in the stereoscope experiments as compared to

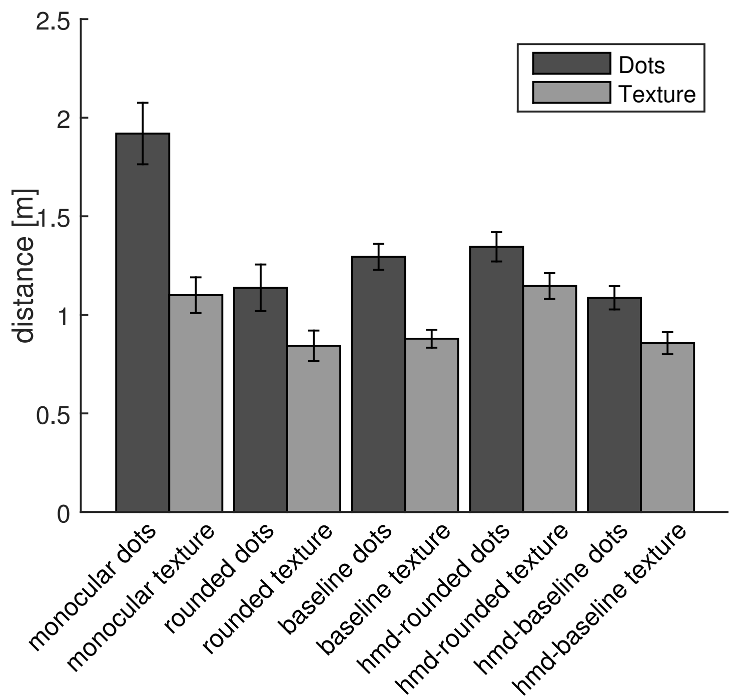


Figure 3.8: Comparison of statistical error for all experiments. From left to right: stereoscope monocular, stereoscope rounded corners, stereoscope baseline, HMD rounded corners, HMD baseline. Light gray=texture condition, dark gray=random-dot condition.

the HMD. This indicates that in the HMD experiments participants were able to perceive the room almost as well from the random-dot stimulus as they could from viewing the textured environment.

The search times for all experiments are compared in figure 3.9. The longest search times are found for the rounded corners on the stereoscope for both conditions. For the texture condition the search times for the baseline stereoscope experiment and the two HMD experiments are comparable. The search times for the random-dot condition are shorter in the HMD experiments compared to the stereoscope. Again it can be seen that the difference between the two conditions was smaller for the HMD experiments.

Figure 3.10 compares the encoding time over all five experiments. For the texture condition the encoding times for the baseline experiment and the to HMD experiment are almost the same. They spent more time encoding the location in the monocular and the rounded corners condition with the stereoscope. In the random-dot condition, the longest encoding time was in the monocular experiment. Here they clearly needed more time to figure out their position than in all other experiments. What is noticeable in the random-dot condition, is the difference between the HMD and the stereoscope. Encoding times were significantly shorter when using the HMD. Especially in the baseline experiment, there was almost no difference in encoding time between the two conditions. That means that participants were almost immediately able to perceive the room when viewing it as a random-dot pattern with the HMD.

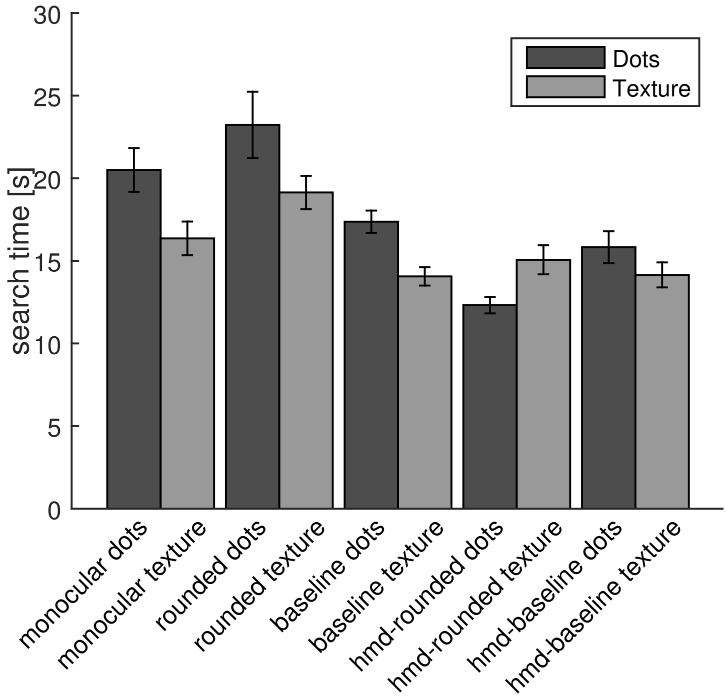


Figure 3.9: Comparison of search time for all experiments. From left to right: stereoscope monocular, stereoscope rounded corners, stereoscope baseline, HMD rounded corners, HMD baseline. Light gray=texture condition, dark gray=random-dot condition.

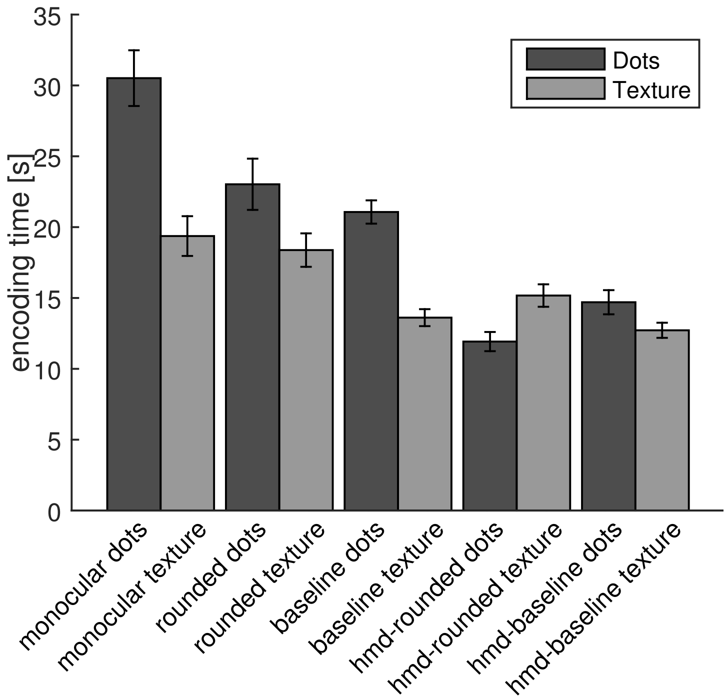


Figure 3.10: Comparison of encoding time for all experiments. From left to right: stereoscope monocular, stereoscope rounded corners, stereoscope baseline, HMD rounded corners, HMD baseline. Light gray=texture condition, dark gray=random-dot condition.

3.4 Discussion

In this chapter two experiments were described that were aimed at replicating the experiments from chapter 2 using a head-mounted display. First, the results of the baseline experiment using the HMD will be discussed, followed by the discussion of the experiment with the rounded corners, and finally the results of the HMD will be compared to the results of the stereoscope.

Baseline HMD

In general it can be said that the results of this experiment match the results of the stereoscope baseline experiment. Participants were clearly able to accomplish the task when they had to rely on depth information only. As in the stereoscope experiment, they made more qualitative errors in the random-dot condition, but this difference to the texture condition was not significant. The statistical error was significantly smaller in the random-dot condition, compared to the stereoscope. Thus participants had little problem in finding the correct segment, but the exact location was easier to find in the texture condition. This shows again that place recognition and navigation based on visual depth perception alone is possible, but that performance gets better as more cues become available.

The distribution of initial walking directions show that searching behavior was clearly goal-directed in both conditions. The fact that there was almost no difference between texture and random-dot suggests that participants could immediately perceive the room in the random-dot condition, and that they had a very good estimate of their own position within this room and the direction of the goal.

Rounded Corners

The differences between the HMD experiments with sharp and rounded corners are comparable to those observed for the respective experiments with the stereoscope. The error rate was not significantly higher using rounded corners showing that they could find the correct segment with equal accuracy. The statistical error was significantly larger than in the room with sharp corners. These results indicate again that place recognition based on depth information only is possible without having landmark information. The vicinity of a memorized place can be found without problems, but the exact estimation of a location gets better when additional cues are available.

HMD vs. Stereoscope

From what is explained in 3.1 we would have expected participants' navigational performance in the random-dot condition to be increased when using the HMD instead of the stereoscope. When looking at the results of the HMD experiment to examine if such an increase is to be found there are two different aspects of performance that have to be considered separately.

The first aspect is the actual navigation and place recognition performance. This performance is measured by participants' error rate, the statistical error and the time they spent searching for the goal. The error rate was lower in the HMD experiment without being significant. This shows that they were able to find the correct segment of the current goal equally well in both experiments. The statistical error, however, was significantly lower in the HMD experiment, i.e. they found the exact goal location with a higher accuracy. This shows that distance estimation from depth perception is enhanced by using the HMD. Also the search time was significantly shorter in the HMD experiment, indicating that it was easier for them to navigate in the random-dot environment.

The second aspect is the self-localization and memorization performance. This refers to participants' ability to estimate and learn their own location within the kite-shaped room. This encoding time was significantly shorter in the HMD experiment, both for the baseline and the rounded corners.

Taken together these findings show that using a HMD greatly increases participants' ability to perceive depth from stereo vision and thus enhances their distance estimation. This better depth-estimates together with the additional proprioceptive feedback and the higher sense of presence increases participants' ability to build up and update a spatial image of their surrounding environment. Therefor it can be concluded that place recognition and navigation in virtual environments greatly benefit from using head-mounted displays.

Chapter 4

Place Recall in a Virtual City

The experiments presented in chapter 2 and 3 showed how locations in a small-scale virtual environment are learned and what is needed to represent a location within such an environment. The experiment in this chapter deals with the navigation in large-scale virtual environments. The first question is how spatial knowledge is transferred from longterm to working memory. A second question that can be asked is how results of this VR experiment relate to results of previous experiments that were carried out in real world environments.

As stated in 1.1.3 when planning a route, representations of the target place are loaded from longterm into working memory. This spatial recall and thus the current representation within our WM may depend on a variety of factors. The factors that we were interested in in this study were the approach direction, aerial direction, and body- or head direction during spatial recall, relative to the location that is recalled from longterm memory. To investigate the influence of each of these factors two different types of memorized places are considered. On the one hand there are places that are lying on the planned route and which are therefor task-relevant. On the other hand there are arbitrary places the are lying off the route and therefor are regarded as being task-irrelevant. The hypothesis is that, based on the concept of a view graph as described in 1.1.3, the representation in WM will differ between these places. In the case of task-relevant locations, participants should load the adjacent view, connected to their current view into WM. Thus the approach direction is expected have a major influence on the recall process. In contrast, for a place that lies off the route and with no distinct approach direction, there is no associated view to be loaded. Thus in these cases the aerial direction and/or the head direction are expected to have a stronger influence.

An experiment was designed where participants could walk freely through a computer simulation of the historic city of Tübingen. The virtual environment that was used for this purpose was the Virtual Tübingen 3D model (Meilinger et al., 2008) that was originally created at the Max Planck Institute (MPI) for Biological Cybernetics. As the model was not "complete", e.g. with the *Stiftskirche* not being included and some streets leading into nowhere, a few modifications had to be added. A detailed description of how the model was built and what has been added is given in 4.1.2. Figure 4.1 shows a bird's-eye view of the original model. The virtual environment was again viewed by wearing the Oculus Rift

Figure 4.1: Bird's-eye view of the original Virtual Tübingen model

DK2 as in chapter 3. Participants then followed an unknown route through Tübingen and were asked to draw sketch maps of well known places within the city. To draw the maps we used a custom-built camera-desk setup that allowed participants to draw their sketch maps within the virtual environment without taking off the HMD. This was important in order to ensure that the immersion and presence was maintained throughout the whole experiment. Therefor participants always had the impression of actually being in the city. Otherwise, if participants had drawn maps with the HMD removed, they would have immediately perceived the lab environment. This would have changed the spatial context in which they had to recall information from longterm memory, making the virtual city a distant location relative to the lab.

4.1 Methods

41 participants, 24 male and 17 female, took part in this experiment. Participation was voluntary and without payment. We recruited participants either by postings or via a university circular email. Subsequent to the experiment participants were asked to fill out a short questionnaire (appendix A.1.1). The mean age was 26.7 years ($\pm$11.6 years) and on

average they have been living in Tübingen for 130 months (±137.5 months). Furthermore the questionnaire revealed a self-assessed local knowledge of 6.5 (±1.6, 1=low, 9=high) and an ability for orientation of 6.1(±2.3, 1=low, 9=high).

4.1.1 Places and Routes

Three places were chosen that participants should draw as sketch maps. Those places had to satisfy the following criteria. First, all three places are frequently visited by many passers-by, so they are well known to many people living in Tübingen. Therefor it was most likely that the participants were familiar with the places and were able to draw them. Second, the places had to have a common name that is known to many people. This was necessary because when asking the participants to draw a map they needed to know right off which place was meant. Otherwise, if the experimenter had to explain or describe the place to them, this would have potentially influenced their drawings. The last constraint was that participants should draw the target places from two different locations on their route and that these locations should lie in different (ideally opposite) aerial directions. Also for both drawing locations it was of course required that participants were not able to see the target place while drawing their sketch map.

Following these constraints we chose the three target places that are highlighted in red in figure 4.2, i.e. the *Krumme Brücke*, the *Marktplatz* and the *Nonnenhaus*. Also shown in this figure is the *Holzmarkt* which was used as target place in previous studies (Röhrich et al., 2014). Although this is a well-known place for most people living in Tübingen we could not use this place, as in the Virtual Tübingen model there is only one way leading towards it. Because of this there were no two possible drawing locations in the virtual environment that differed in their aerial direction relative to the *Holzmarkt*. The route that participants had to walk is also depicted in figure 4.2. There were two different starting points for the route, one in the west in the Haaggasse and one in the east at the *Holzmarkt*.

The two different route directions (starting east vs. starting west) and the two different drawing locations (henceforth referred to as A and B) for each target place made up for a total of four different experimental conditions. As participants could only draw a specific target place once, we divided them into four corresponding groups. Thus each group began their route either from the western or the eastern starting point and then drew their sketch maps either at location A or B for the respective target place. The resulting target place vs. drawing location constellation was as follows.

- *Krumme Brücke*: the *Krumme Brücke* was the only place that was actually lying on the route, i.e. being visited by the participants. The drawing locations were at the Neugäßle (B) and at the Café Hirsch (A). That means depending on the starting point, participants drew the sketch map either before or after crossing the *Krumme Brücke.*

- *Marktplatz*: despite being in the vicinity of the traveled route, the *Marktplatz* was not visited by the participants and it was also not visible for them at any time. The drawing locations were at the Neugäßle (A) and the Holzmarkt(B).

- *Nonnenhaus*: the *Nonnenhaus* was further away compared to the other two locations and was virtually outside the region of the experiment. It is therefor more comparable to the distance condition from *Röhrich et al.*. The drawing locations were at the Café Hirsch (B) and the Holzmarkt (A).

The following table gives an overview of the four different test groups. For each group the respective starting point and for each target place the corresponding drawing locations for this group are listed. That means, for example, that participants in test group A-East

	Test Group			
	A-East	A-West	B-East	B-West
Starting Point	East	West	East	West
Krumme Brücke	A (Café Hirsch)	A (Café Hirsch)	B (Neugäßle)	B (Neugäßle)
Marktplatz	A (Neugäßle)	A (Neugäßle)	B (Holzmarkt)	B (Holzmarkt)
Nonnenhaus	A (Holzmarkt)	A (Holzmarkt)	B (Café Hirsch)	B (Café Hirsch)

Table 4.1: Combinations of start conditions and drawing locations for each target place.

were starting east at the Holzmarkt and first drew the *Nonnenhaus* at the Holzmarkt, then the *Krumme Brücke* at the Café Hirsch and finally the *Marktplatz* at the Neugäßle. Participants were randomly assigned to one of the four test groups.

4.1.2 Setup

Apparatus

Participants should be able to walk and look around freely in the computer simulated City of Tübingen. Therefor they were again wearing the Oculus Rift as in the previous studies, using the same hardware setup. Also the navigation was identical, that is they used the game controller for translations and controlled their movement direction by turning the head.

As described above, the drawing of the sketch maps was a bit complicated because participants could not just be asked to take off the HMD and draw on a piece of paper, as

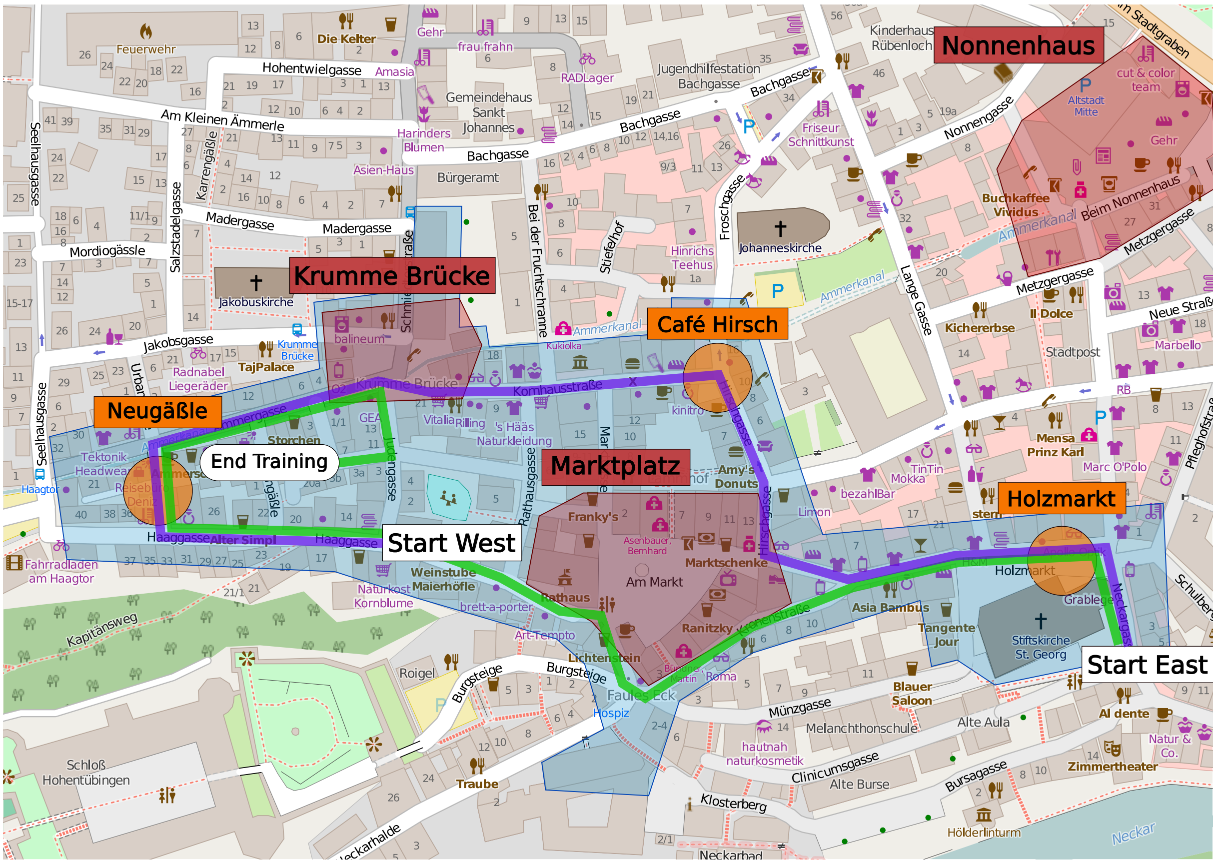
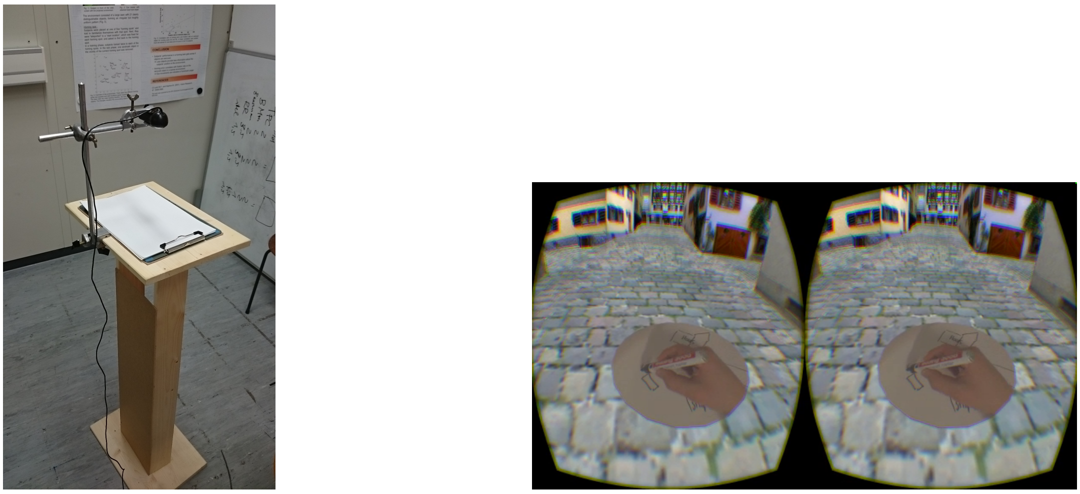


Figure 4.2: Map of Tübingen (OpenStreetMap©). Blue = area covered by the Virtual Tübingen model. Red = target places that were drawn by the participants. Orange = drawing locations. Green line = training route. Purple line = test route with respective start positions.

this would have disturbed the whole spatial context. To prevent them from being taken out of the environment, we constructed a simple desk, with an height of 1.05m and a logitech® QuickCam® Pro 9000 webcam attached to it. On the desk was a clipboard with a stack of paper and the camera position was adjusted so that it "looked down" on the paper. Figure 4.3a shows this setup. Whenever participants had to draw a map, the desk was placed in front of them by the experimenter. The live video from the camera was then displayed right in front of them inside the virtual environment. Thus when they put their hands on the desk underneath the camera, the could see their own hand movement in virtual reality. Provided with a felt pen and after a short training, participants were then able to draw their sketch maps while still being in the simulation. An example of participants' view of their hands in virtual reality when standing in front of the desk is shown in figure 4.3b. As can be seen in this figure, the input from the camera was displayed as a round sheet of paper in the simulation. This shape was used to prevent participants from being biased to draw maps oriented in one of the cardinal directions.

To let the experimenter see what the participants saw and what they did in the simulation, the display from the HMD had to be streamed to the main monitor. This was



(a) custom built desk (b) hands in vr

Figure 4.3: Webcam setup for displaying participants' hand in VR. (a) Custom built desk with webcam attached and clipboard. (b) View of participants' hands in VR while drawing.

achieved by using a program called Open Broadcaster Software (OBS) that is normally used to stream live videos to online services such as Youtube™. This was not only useful information while controlling the experiment, but was also essential for the experimenter to give verbal instructions in the training-phase as described in 4.1.3. As a side effect, this also provided the possibility to record videos from each trial.

Virtual Environment

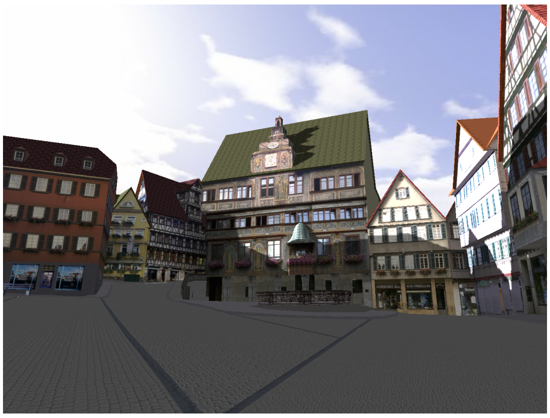
The virtual environment that was used was mostly the same Virtual Tübingen as used by Meilinger et al. (2008), a detailed description of the model and the building process can be found at *http://virtual.tuebingen.mpg.de/index.html* (MPI for Biological Cybernetics, 2005). Building the model basically consisted of two steps.

In the first step, the geometry of the buildings was constructed using architectural drawings. These drawings were digitized using a drawing tablet and then polygonal 3D-models were created from these digitized drawings using a custom-made software tool. Finally the models were loaded into a 3D modeling program and were improved and corrected.

In the second step the textures were created. For this process, photos had to be taken for each building that has been modeled. The images were then corrected perspectively and for differences in lighting and shadowing and finally placed on the 3D models. Figure 4.4 shows an example view of the *Marktplatz* as it can be seen in the final model. Also shown for comparison is a real photograph of the same scene, taken from a similar perspective.

As already mentioned above, the model that we received from the MPI was partially

(a) *Markplatz* photograph



(b) *Markplatz* VR

Figure 4.4: Example view of the virtual Tübingen model showing the *Marktplatz* (b) compared to a real photograph of the same scene (a), taken from a similar perspective

incomplete. From figure 4.1 it can be seen that in the eastern part of the model the *Holzmarkt* is included, but a 3D model of the *Stiftskirche* is missing. This is a very prominent building at this place, and not having this in the simulation not only leaves an empty space but also changes the whole appearance of the *Holzmarkt*. Therefor a new 3D model had to be built. As we did not have any architectural drawings or construction plans of the *Stiftskirche*, the 3D reconstruction is not accurate regarding scale and proportion of the church. The only point of reference that could be used for constructing the model was the groundplan. In the original model, at the location where the *Stiftskirche* was supposed to be, the contour of the building's layout was cut out so that the groundplan could be inferred from this contour. A polygonal 3D-model was then built using Multigen Creator from Multigen-Paradigm®, a program for creating realtime 3D models. The height of the building was estimated from comparing the neighboring houses included in the model with photos from the *Holzmarkt* showing the *Stiftskirche* and the neighboring houses. In the next step photos were taken from the *Stiftskirche*. These photos were perspectively corrected and applied as textures to the 3D model. The result can be seen in figure 4.5. Again a real photo of the same scene taken from a similar perspective is shown for comparison.

Another problem with the original model that can be seen in figure 4.1 was that there were some streets and lanes around the *Holzmarkt* that led into nowhere. This means participants would have seen where the simulation ended and in the worst case might have fallen off the edge of the world. To avoid this we closed these gaps by placing a simple plane where the model ended. Then photos were taken from a perspective viewing down the respective streets and these photos were applied as textures to the planes. A similar method was used by the MPI at other locations. Figure 4.6 shows two such streets with an open end compared to a closed view. When standing away far enough, the illusion of a continuous world was almost perfect and only when coming too close to such a "wall"

(a) *Stiftskirche* photograph (b) *Stiftskirche* VR

Figure 4.5: Example view of the virtual Tübingen model showing the *Stiftskirche* (b) compared to a real photograph of the same scene (a), taken from a similar perspective

participants would notice the delusion.

One final change to the model was added by Nils Brehm (Brehm, 2015) to obscure the view of the *Marktplatz*. This was necessary, because the *Marktplatz* was lying off the route and should not be seen by the participants. However, there was a small spot along the route where participants could have seen the place. To prevent this, the view was blocked by placing a 3D modeled truck between the route and the *Marktplatz* as shown in figure 4.7.

(a) open streets (b) closed streets

Figure 4.6: (a) Example of an open street where the end of the world was visible. (b) The same scene with the streets closed by textured planes.

Software

The experiment was programmed using Unity 4, a cross-platform game engine developed by Unity Technologies, which is widely used to develop games for different platforms. To integrate the Oculus Rift into our experimental software, the Unity 4 Integration V0.4.4-beta provided by Oculus VR was used together with the Oculus SDK for Windows V0.4.4-beta. This integration package offers a first person controller that includes all necessary features, such as head tracking or distortion correction.

(a) *Marktplatz* visible (b) *Marktplatz* view blocked

Figure 4.7: Preventing participants from seeing the *Marktplatz*. (a) *Marktplatz* visible from the route. (b) *Marktplatz* view blocked by truck.

One important aspect in developing the software was to keep the frame rate constantly high. This was crucial when using the Oculus Rift. A drop in frames per second (fps) or an increased latency would be perceived as unpleasant by the participants and increases the chance of feeling sick. To increase performance in this rather large and detailed environment, participants' view was limited to 300 meters. Because of the course of the roads in the virtual city participants could not notice this limitation at any time. Another way to gain more performance was to deactivate realtime shadows. For some reason, pre-computing the shadows did not work, at least in the free version of Unity used here. Thus completely disabling shadows reduced the realism of the scenery but this trade-off was acceptable to keep the performance high.

To implement the drawing in virtual reality a Unity feature called *webcamTexture* was used. This *webcamTexture* provides an interface to the array of webcams connected to the computer. The video input from the attached webcam can then be applied just like a normal texture to any surface within the virtual environment. A virtual clipboard displaying this *webcamTexture* was then positioned within the scene in such a way that it was always displayed in front of the participants when they were stopped for drawing.

4.1.3 Procedure

Participants were informed about the experiment itself, the collected data and the potential motion sickness caused by wearing the HMD. They gave written consent that they had read and understood the terms of participation. Subsequently they were given printed information of the experimental procedure and what the task was about. They were also verbally instructed by the experimenter and occurring questions were answered.

After being instructed, they put on the HMD and the experimenter helped them to adjust it so that they felt comfortable wearing the device. Participants were standing throughout the whole experiment, except for two older participants who had to sit down to complete the task because of motion sickness and vestibular disorder. Finally they were given the game controller and the simulation was started.

To familiarize participants with the HMD and the navigation in the virtual environment they had to complete a training route (figure 4.2, green route). At the beginning of this training route they were standing at the *Holzmarkt* next to the *Stiftskirche*. They were then verbally instructed by the experimenter where to go, e.g., "go straight forward and then turn left". When they reached the end point of the training route, a small side-street called Süßenloch, they were stopped and could no longer move around. However, they could still look around freely by turning their head. The camera image was then displayed in front of them and the experimenter placed the desk as described above. They were then instructed to draw a sketch map of the *Neue Aula*, a place that is also well known to many people living in Tübingen, especially students. Drawing this map in the training-phase should help participants to get used to the drawing in virtual reality. Although the scale of the hands displayed as well the delay of the perceived movement made this way of drawing feel a bit strange at the first moment, participants adapted quite quickly. After this short training they were able to draw almost naturally and produced nice evaluable maps of the target places.

When they had finished their practice drawing and felt comfortable enough with this setup they were teleported to one of the two possible start positions (east or west) and the main task began. To show participants the route they had to travel a large three-dimensional arrow was displayed whenever they had to change their walking direction. If no such information was displayed participants were instructed to walk straight ahead and follow the course of the road until the next arrow appeared. Whenever they arrived at one of the three drawing locations they were again stopped as in the training-phase and were placed in front of the drawing desk. The experimenter verbally instructed them to draw one of the three target locations depending on the current drawing location and experimental condition (A or B). For example, they were told, "Please draw a sketch map of the Marktplatz". After completing their drawing they put the pen aside and the desk was removed. The experimenter then unlocked the movement again by pressing F1 on the keyboard and participants continued their route to the next drawing location. Finally they arrived at the goal position and the experiment ended. Subsequently participants were given the questionnaire to fill out.

4.2 Results

4.2.1 Participants

Table 4.2 shows the evaluated answers from the questionnaire separately for each group and for all groups together. There were no significant differences between the four groups regarding age, orientation ability and local knowledge.

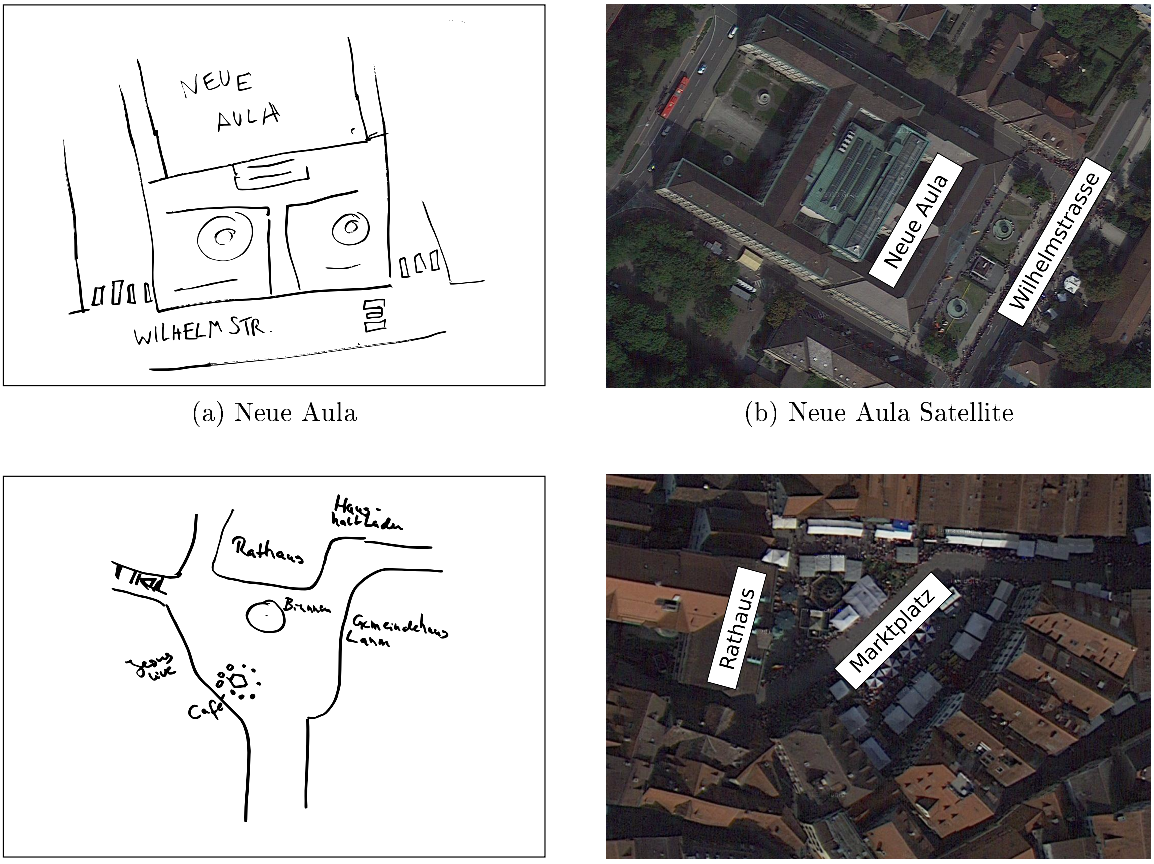
	age (years)	period of residence (months)	male	female	orientation ability	local knowledge	difficulty	fun
A-East	26.4 ±10.9	170.8 ±147.5	8	3	7.3 ±0.8	7.1 ±1.1	6.9 ±1.2	7.9 ±1.5
A-West	23.6 ±7.8	51.1 ±82.4	4	6	4.6 ±2.5	5.6 ±1.7	5.5 ±1.6	8.5 ±0.5
B-East	26.9 ±13.7	165.5 ±145.2	6	4	6.4 ±2.2	6.6 ±1.6	6.3 ±1.9	8.5 ±0.7
B-West	30.1 ±16.4	128.5 ±155.2	6	4	5.9 ±2.6	6.5 ±1.9	6.6 ±1.2	8.6 ±0.5
total	26.7 ±11.6	130 ±137.5	24	17	6.1 ±2.3	6.5 ±1.6	6.3 ±1.5	8.4 ±0.9

Table 4.2: Evaluated answers for each of the four test groups and for all groups together

4.2.2 Map Orientations

As explained above each participant drew one sketch map as an exercise example at the end of the training-phase and three sketch maps during the test-phase, making up for total of 164 drawings to be evaluated. The evaluation procedure was as follows.

First, all maps were rated by three independent raters. As the practice drawings of the *Neue Aula* were not of interest for later analysis, these maps were only rated by two persons. For the rating, the maps were assigned one of eight compass directions (north, northeast, east, etc.) by comparing each map to a north-oriented reference image of the respective place, taken from Google Maps©. The reference image was then rotated to match the sketch map's orientation and the upward facing compass direction of the reference was chosen as the drawing's orientation. Figure 4.8 illustrates this for two example maps from the *Neue Aula* and the *Marktplatz*. The reference image shown in figure 4.8b for example had to be rotated 45°clockwise to match the orientation of the sketch map in figure 4.8a. Thus the rotated reference image was northwest-oriented and this orientation was assigned to the sketch map. Likewise the same procedure yielded a west orientation for the drawing of the *Marktplatz* shown in figure 4.8c. Note that here in this example figure the reference images are shown as satellite images for better illustration whereas in the actual rating procedure street maps were used. This rating resulted in three orientations for each sketch map and the final orientation was chosen as



(a) Neue Aula (b) Neue Aula Satellite

(c) Marktplatz (d) Markplatz Satellite

Figure 4.8: Example sketch maps illustrating the rating procedure. First row: Sketch map of the *Neue Aula* (a) with respective satellite view (b) showing north west orientation. Second row: Sketch map of the *Marktplatz* (c) with respective satellite view (d) showing west orientation.

the compass direction where two out of three raters agreed in their judgment. Following this criterion 99% (122 out of 123) of the sketch maps were included in the subsequent analysis, whereas a total agreement of all raters was found in 60% of the cases. As the practice drawings were rated by two persons, only those were further analyzed for which both of the raters agreed in their judgment, resulting in a total of 85% (35 out of 41).

The rated maps of each target place were then sorted according to the four experimental conditions as shown in table 4.1 and the orientations were converted into degrees (east $\widehat{=}$ 0°, northeast $\widehat{=}$ 45°, north $\widehat{=}$ 90°, etc.) for further calculations. The distribution of orientations were plotted as circular histograms in the same way as the initial walking directions in 2.2 and 3.3, results are shown in figure 4.9 - 4.12. Here the full circle was divided into bins of 45°and labeled with the respective compass directions. Again the mean resultant vector $\vec{r}$ was plotted showing how concentrated the orientations were lying around a specific drawing direction. All statistics were done using Matlab 2015a© and the *Circular Statistics Toolbox* (Berens, 2009) as in the previous sections.

Factors Influencing Map Orientations

There are several factors we looked at that might have influenced spatial recall and thus have an effect on the orientation of the sketch maps. These factors were considered to be the approach-, aerial-, and the head direction.

The aerial direction is the angle between participants' current drawing location and the center of the respective target place. This was measured by calculating the mean drawing locations over all participants and marking those coordinates on a map of Tübingen. Then the angle between those markers and the center of the target place was measured with an on-screen triangle ruler.

The approach direction is defined as the direction in which participants were oriented when walking towards the respective target place along their route. This factor was only taken into account for the *Krumme Brücke*, as this was the only target place that participants visited along their route. We measured this direction by looking at the final segment of the route leading towards the *Krumme Brücke* and taking the angle between the starting point of this segment and the center of the *Krumme Brücke*.

Participant's head directions were taken from the recorded log files and are plotted as circular histograms together with the histograms of the map orientations at the respective drawing location (figure 4.9-4.12). Note that participants were able to look around freely while drawing the map. Nonetheless it seems that they rather kept their eyes on the virtual clipboard in front of them than looking around for orientation. Looking at the recorded head directions during drawing shows a mean range of 138.7°in the training-phase and ranges between 73°and 99.1°in the test-phase. As these directions were recorded from the moment participants were stopped to the moment the experimenter continued the trial, head movements before and after the actual drawing are also included. Thus it is not possible to tell where participants looked during drawing relying solely on the logged data. One additional information about head directions came from the recorded videos, showing participants' point of view during drawing. It can be clearly seen that they kept looking at the clipboard and that the larger changes in head direction occurred before or after drawing. For example one participant had trouble to find the clipboard in virtual reality, turning his head around the whole 360°, whereas another one looked around in a wider range after having finished the map. We therefor assumed the head direction recorded the moment they were stopped before the drawing to be the participants' head direction during spatial recall in the drawing task. Of course, in humans the viewing direction is up to some degree independent of our head direction. We did not measure viewing direction directly, due to the missing possibility of using eye tracking together with the Oculus Rift. Thus we assumed the viewing direction to be identical to the head direction in first approximation.

Neue Aula

Figure 4.9 shows the results of the sketch maps from the *Neue Aula*. Participants drew the *Neue Aula* at the end of the training-phase, as an exercise example. Therefor this map was drawn by all participants at the same location and under the same conditions independent of which group they were assigned to in the test-phase. Thus there is only one circular histogram containing all 35 maps that passed the rating procedure.

The *Neue Aula* is located on the Wilhelmstrasse and is oriented along the northwest-southeast axis, whereas the road runs along the northeast-southwest axis. However it can be seen that the preferred drawing direction for the *Neue Aula* was northwest (23 out of 35 maps). That means that most of the participants drew the Wilhelmstrasse horizontally with the *Neue Aula* above the road oriented along the north-south axis.

A rayleigh test for non-uniformity of circular data showed that the map orientations are non-uniformly distributed ($p<0.001$), i.e. participants showed a clear preference for one map orientation. The mean orientation over all participants was 127.31°, corresponding to the northwest direction mentioned above. The aerial direction and participants' head direction during drawing were 34°and 195.7459°respectively. A two-sample kuiper-test showed a highly significant difference between the mean map orientation and the head direction ($p=0.001$), indicating that the head direction had no influence on the maps. To test for the influence of the aerial direction a circular v-test was used. Results show that the distribution of map orientations was not centered around the aerial direction ($p>0.5$).

Krumme Brücke

The results of the sketch maps drawn of the *Krumme Brücke* are shown in figure 4.10. The *Krumme Brücke* is oriented along the west-east axis and participants also crossed this place along this axis either from west to east or from east to west. Drawing locations were at the *Neugässle* and the *Café Hirsch*. The different directions explained in 4.2.2 were determined for each of the four experimental conditions, results are summarized in table 4.3.

A rayleigh-test showed a non uniform distribution for all conditions, i.e. participants seemed to have a preferred drawing direction. It can be seen that this preferred drawing direction differed between the conditions. For the west conditions the maps were oriented in the east or north-east direction (A-West=30.94°, B-West=27.02°), whereas in the east condition the dominant orientation was west or north-west (A-East=135°, B-East=152.14°). We then tested for the influence of the different factors on the map orientations. The detailed results of the tests are shown in table 4.3 together with the re-

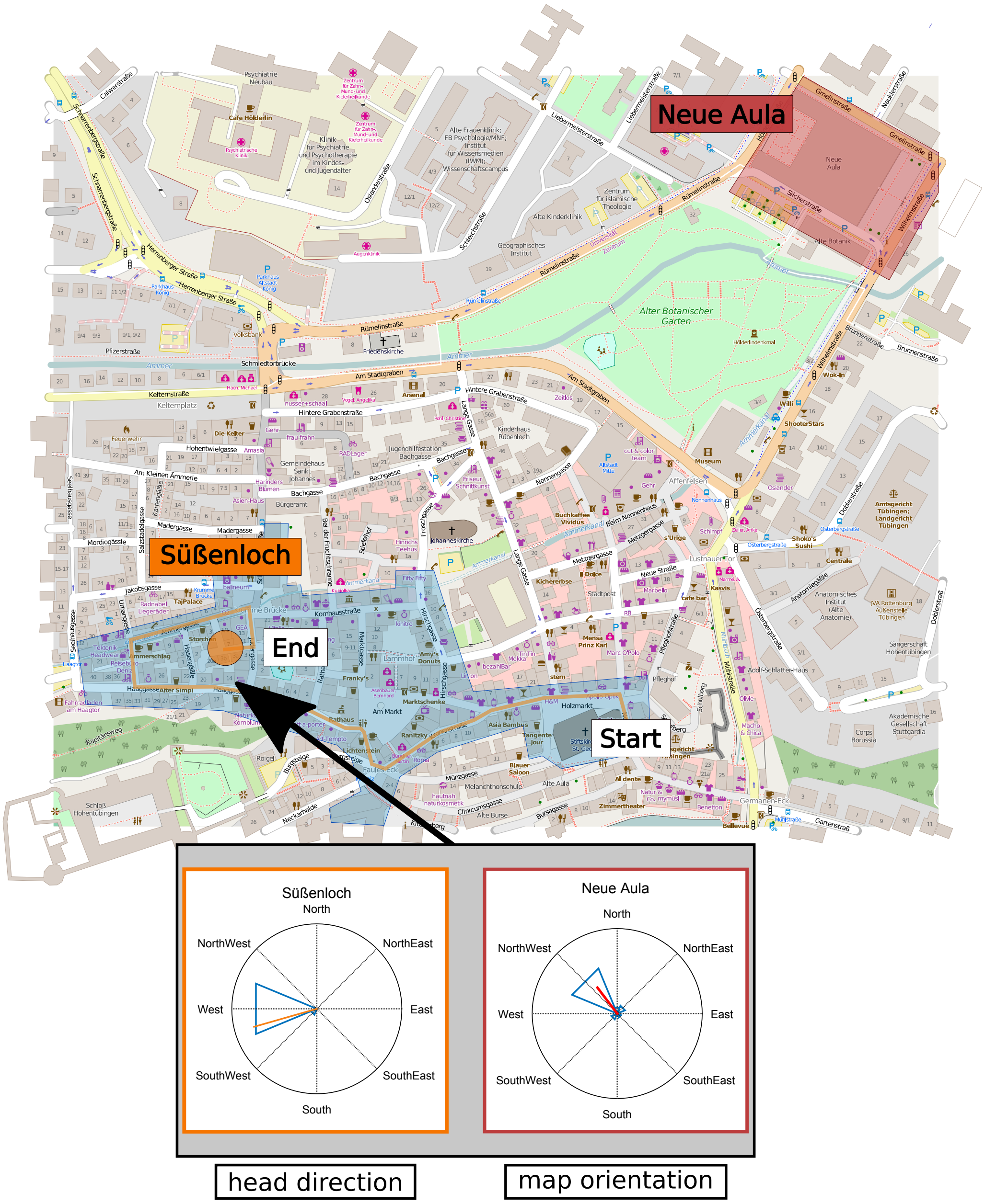


Figure 4.9: Results of the *Neue Aula*. The red area shows the target place, the drawing location is depicted as an orange circle. Also shown is the training route with corresponding start and end location. Red Square: circular histogram of the distribution of map orientations over all participants, red line depicting the mean resultant vector. Orange square: circular histogram of the distribution of head directions. For comparison all histograms have been normalized to the total number of participants per drawing location (N=35).

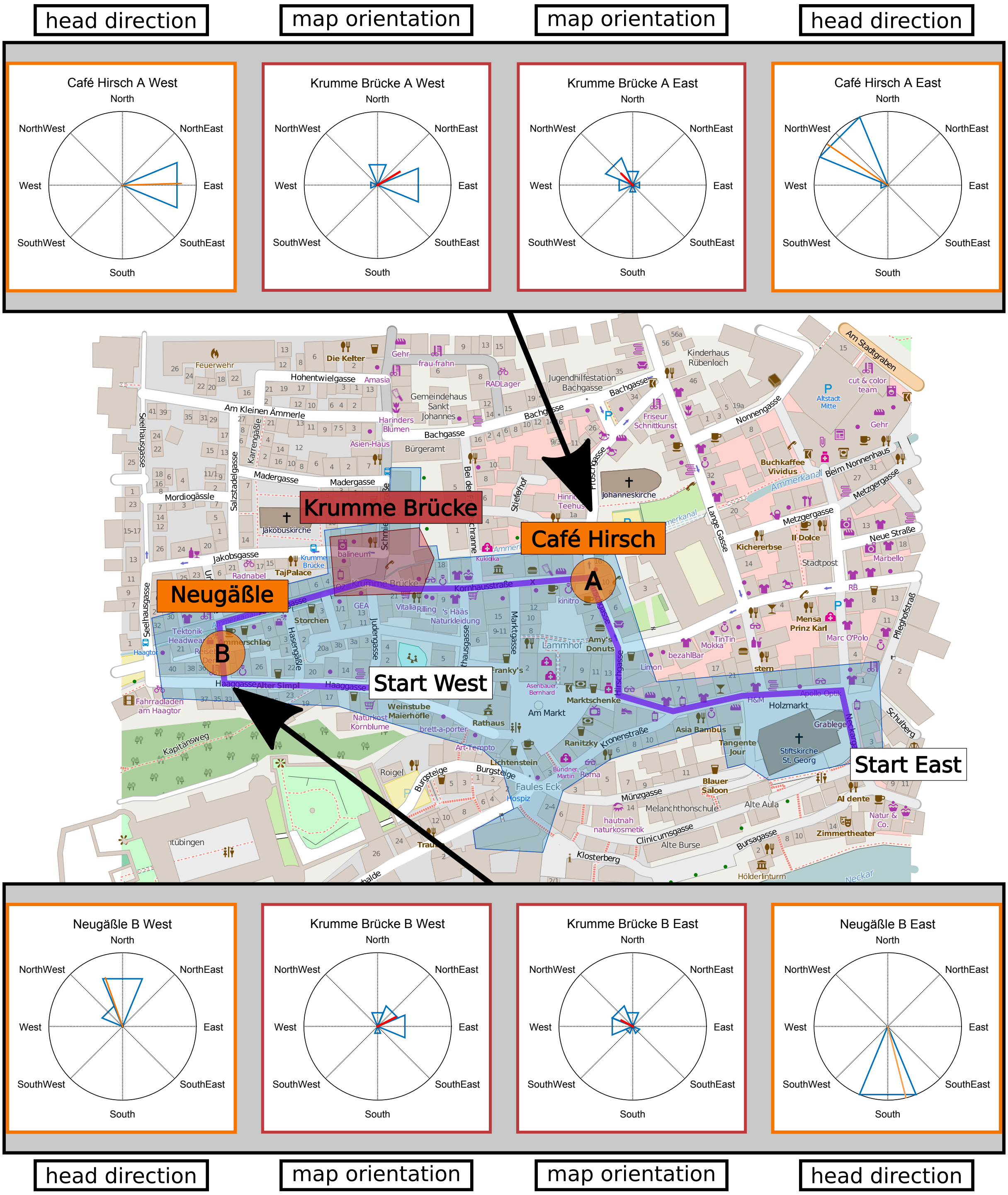


Figure 4.10: Results of the *Krumme Brücke*. The red area shows the target place, the drawing locations are depicted as orange circles. Also shown is the test route with corresponding start and end location. Red Squares: circular histograms of the distribution of map orientations over all participants, red line depicting the mean resultant vector. Orange square: respective circular histograms of the distribution of head directions. For comparison all histograms have been normalized to the total number of participants per drawing location (N=10).

spective factor. A two-sample kuiper-test showed a significant difference between the map orientation and the head direction in 3 out of 4 conditions. Only in condition A-West did the two directions not differ significantly ($p>0.1$). For the aerial direction a v-test showed a significant centering of the map orientations around the aerial direction only for A-East and B-West. These are the two cases where the aerial direction is close to the approach direction. In the other two conditions, A-West and B-East, where the aerial direction is almost opposite to the approach direction, there is no such effect. Here the centering around the aerial direction is highly non-significant. Likewise we looked if there was a significant centering of the map directions around the approach direction. Here a v-test showed a significant effect for all four conditions. Thus the approach directions seemed to be the only factor that had an influence on the map orientations in all conditions. This indicates that for places that lie on one's route and which are therefor task-relevant, the spatial recall depends on the view that one would expect to see when approaching this place.

	map orientation	head direction	aerial direction	approach direction
A-East	135°($p<0.05$)	145.97°($p=0.01$)	178°($p<0.05$)	170°($p<0.05$)
A-West	30.94°($p<0.05$)	1.71°($p>0.1$)	176°($p>0.5$)	20°($p<0.01$)
B-East	152.14°($p<0.05$)	284.17°($p=0.001$)	22°($p>0.5$)	170°($p<0.01$)
B-West	27.02°($p<0.01$)	109.48°($p=0.001$)	30°($p<0.01$)	20°($p<0.01$)

Table 4.3: Map orientations for the *Krumme Brücke* for each experimental condition with respective head, aerial, and approach direction. Also shown are p-values for the respective tests: map orientation=rayleigh-test, head direction=kuiper-test, aerial and approach direction=v-test.

Marktplatz

Figure 4.11 shows the results of the sketch maps drawn for the *Marktplatz*. Being a central place in the historic city of Tübingen the *Marktplatz* itself does not show an orientation in one of the compass directions. There is one noticeable building, however, namely the town hall, situated in the west of the place. Also there is a fountain in front of the town hall that is well known to many people. The other buildings surrounding the *Marktplatz* are mostly shops and restaurants that are less salient. As participants did not cross this place on their route, there was no approach direction that might have influenced the drawings. The only condition, where one could argue that the approach direction might be a possible explanation is B-East, as will be explained below. The results of the different directions and the respective statistical tests are shown in table 4.4.

For the two target places mentioned above, the *Neue Aula* and the *Krumme Brücke*,

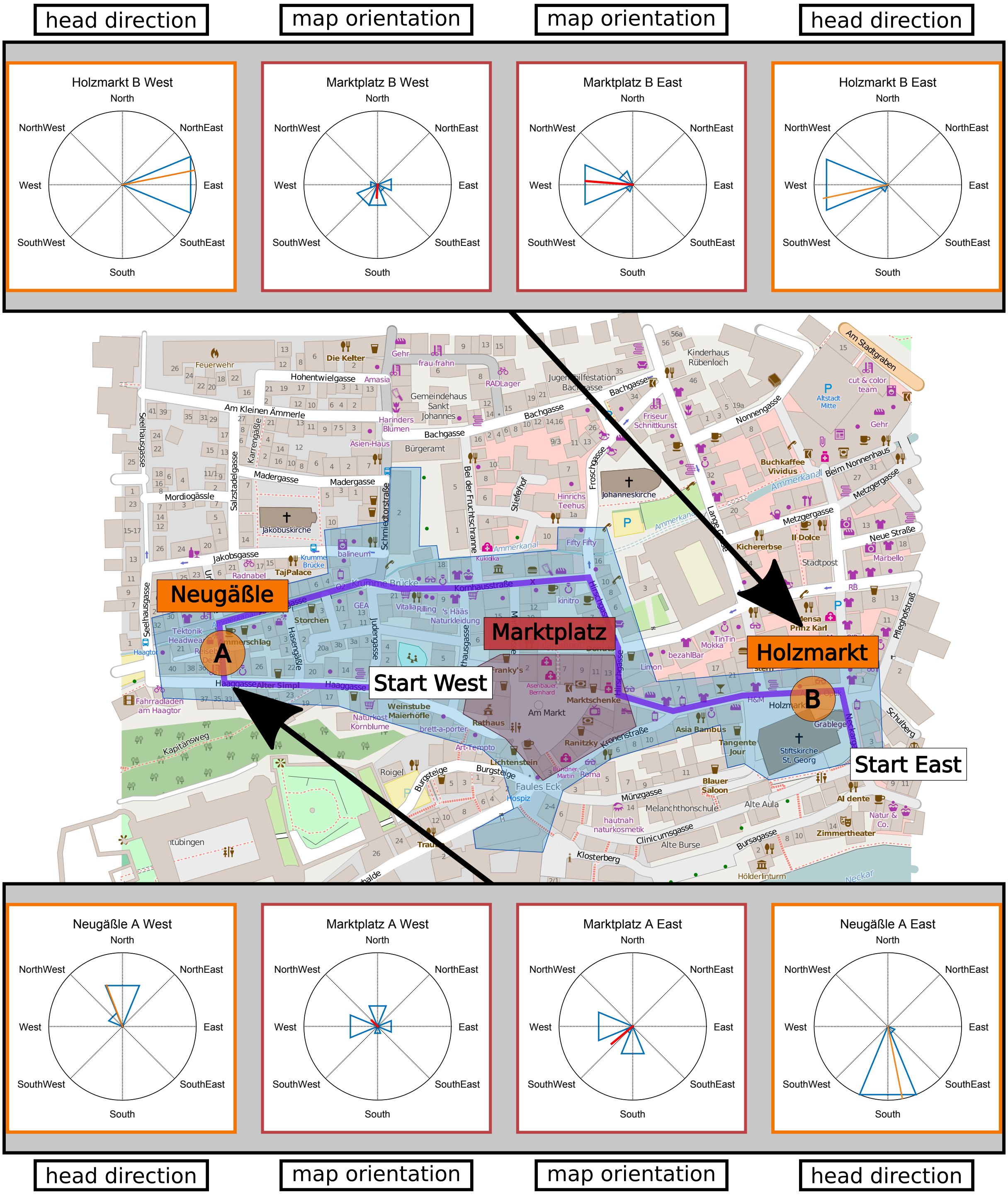


Figure 4.11: Results of the *Marktplatz*. The red area shows the target place, the drawing locations are depicted as orange circles. Also shown is the test route with corresponding start and end location. Red Squares: circular histograms of the distribution of map orientations over all participants, red line depicting the mean resultant vector. Orange square: respective circular histograms of the distribution of head directions. For comparison all histograms have been normalized to the total number of participants per drawing location (N=10).

there was a preference in the drawings for one direction depending on the condition. However, for the *Marktplatz* the situation appears to be a bit different. A rayleigh-test showed a non-uniform distribution for 3 out of 4 conditions. The exception is condition A-West ($p>0.1$), i.e. for this condition there was no preferred map orientation. Also for condition A-East there is no such preferred orientation. Here the two dominant drawing directions are orthogonal to each other (west=5, south=4), so although the mean vector clearly points southwest, this is obviously not the preferred drawing direction. Therefor it can be said that for the sketch maps drawn at the *Neugässle* there was no influence of one of the factors.

When looking at the histogram of condition B-West it can be seen that there is an orientation of the sketch maps in south and southwest direction. However, a two-sample kuiper-test showed a significant difference between the map orientation and the head direction. Also, a v-test did not show a significant centering of the map orientation around the aerial direction. Thus none of the factors analyzed here seemed to have an influence on the drawings.

The results of B-East exhibit a clear preference for the west direction. A kuiper-test again did show a significant difference between the map orientation and the head direction, indicating that this factor did not influence participants' drawing. But, in contrast to the other *Marktplatz* conditions, a v-test showed that the map orientations were significantly centered around the aerial direction. This factor might therefor explain the preferred drawing direction. Another influence in this condition might come from the approach direction. Although participants did not cross the *Marktplatz*, they were standing at the *Holzmarkt* facing westwards when drawing the map. Thus they might have anticipated that following the route would lead them towards the *Marktplatz*, just like they would have before walking towards the *Krumme Brücke*. Indeed a v-test revealed that the map orientations are centered significantly around the approach direction (165°, $p<0.1$). As the map orientation (175.75°) lies between the head and the aerial direction (184°and 165°respectively) it is not possible to distinguish the influence of these two factors.

	map orientation	**head direction**	**aerial direction**
A-East	219.52°($p<0.01$)	281.37°($p=0.01$)	347°($p>0.5$)
A-West	135°($p>0.1$)	111.1°($p=0.001$)	351°($p>0.5$)
B-East	175.57°($p<0.001$)	191.22°($p=0.005$)	184°($p<0.001$)
B-West	265.93°($p<0.05$)	11.4°($p=0.001$)	185°($p>0.1$)

Table 4.4: Map orientations for the *Marktplatz* for each experimental condition with respective head, aerial, and approach direction. Also shown are p-values for the respective tests: map orientation=rayleigh-test, head direction=kuiper-test, aerial and approach direction=v-test.

Nonnenhaus

The results of the sketch maps drawn of the *Nonnenhaus* are shown in figure 4.12. The *Nonnenhaus* is a famous shopping center in Tübingen and is oriented along the east-northeast axis. There is also a small river, the Ammer, which runs along the same orientation. It is also notable that the *Nonnenhaus* is situated outside the area covered by the Virtual Tübingen, a bit farther away from the route. Thus participants did not cross this place when following their route. The results of the different directions and the respective statistical tests are shown in table 4.5.

A rayleigh test showed a non uniform distribution for 3 out of 4 conditions, except for B-West ($p>0.1$). However, it can be seen that for both west-conditions there was a preferred map orientation in the east and northeast direction with a very clear northeast orientation for A-West. A two-sample kuiper-test showed a significant difference between the map orientation and the head direction for both west-conditions. On the other hand a v-test showed that for both cases the map orientations are significantly centered around the aerial direction. This seems to indicate that there was rather an effect of the aerial direction than of the head direction.

For the east-conditions there only seems to exist a preferred map orientation in condition B-East and not in condition A-East. In the latter the most dominant orientations are orthogonal to each other (north=3, west=3) with a smaller bin in between (northwest=2). Thus in this condition there seemed to be no influence of one of the factors. In the case of the B-East condition there was a preferred map orientation in the northwest and north direction. Here a kuiper-test showed no significant difference between the map orientation and the head direction whereas a v-test did not show a significant centering of the map orientation around the aerial direction. Therefor, in contrast to the west-condition, its seems here to be an effect of the head- rather than the aerial direction.

	map orientation	head direction	aerial direction
A-East	135°($p<0.05$)	182.94°($p=0.05$)	85°($p<0.05$)
A-West	36.46°($p<0.001$)	12.51°($p=0.05$)	80°($p=0.001$)
B-East	123.3°($p<0.01$)	154.15°($p=0.1$)	26°($p>0.5$)
B-West	22.5°($p>0.1$)	10.2°($p=0.02$)	24°($p<0.05$)

Table 4.5: Map orientations for the *Nonnenhaus* for each experimental condition with respective head, aerial, and approach direction. Also shown are p-values for the respective tests: map orientation=rayleigh-test, head direction=kuiper-test, aerial- and approach direction=v-test.

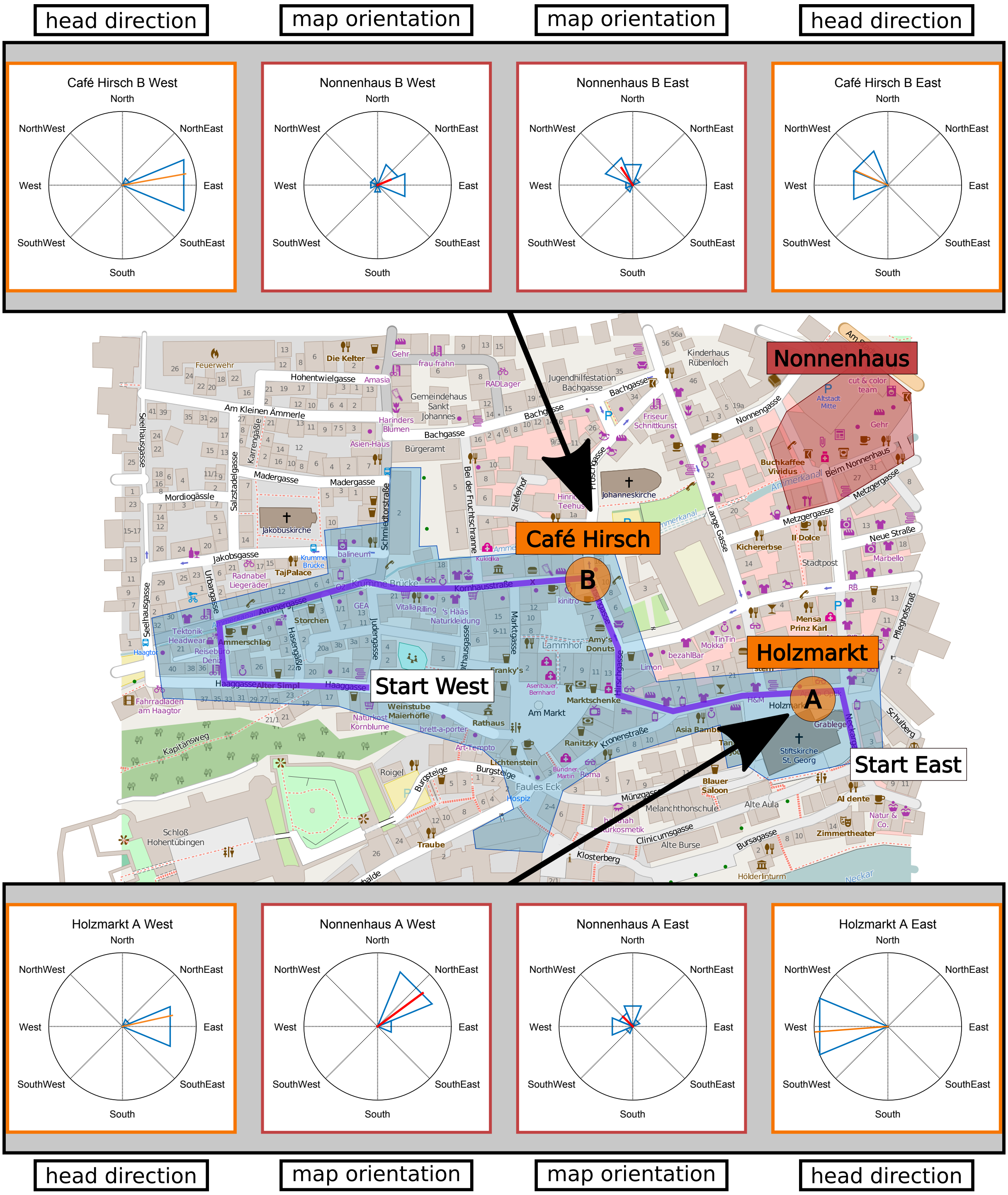


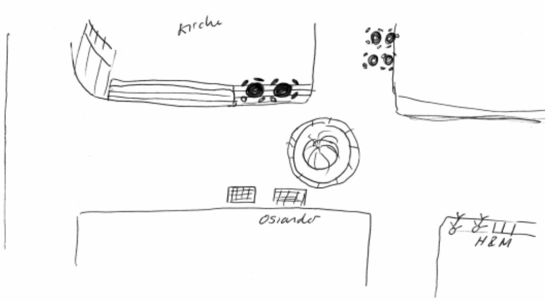
Figure 4.12: Results of the *Nonnenhaus*. The red area shows the target place, the drawing locations are depicted as orange circles. Also shown is the test route with corresponding start and end location. Red Squares: circular histograms of the distribution of map orientations over all participants, red line depicting the mean resultant vector. Orange square: respective circular histograms of the distribution of head directions. For comparison all histograms have been normalized to the total number of participants per drawing location (N=10).

4.3 Discussion

We tested 41 participants in a virtual environment representing the historic city of Tübingen and asked them to draw sketch maps of well-known places while traveling along a given route.

First, we looked at the results of the sketch maps that were drawn of the *Neue Aula* at the end of the training-phase. These maps were meant as an exercise for participants to get familiar with drawing in virtual reality and are not analyzed in detail. However there are some conclusions that can be drawn from looking at these drawings.

The first thing that is to be mentioned is that participants were indeed able to adapt to this way of drawing very well. Of course, the presentation of their hands in VR might have felt a bit strange and unnatural for them in the beginning. This was due to the angle in which they looked at their hands being slightly different from reality, as well as the scale of their hands. Also there was a small input lag which was noticeable while drawing. Nevertheless, with a little practice, participants were able to produce sketch maps that did not differ qualitatively from maps drawn by participants in previous studies. Figure 4.13 shows a comparison between two example drawings. This might not seem such an important aspect at first, but it is crucial if the results presented here are to be compared to previous results.



(a) *Holzmarkt* previous studies

(b) *Neue Aula* current study

Figure 4.13: Comparison of sketch maps from previous studies and the current VR study. Drawings were qualitatively comparable (a) *Holzmarkt* drawn in previous study (Binder, 2012) (b) *Neue Aula* drawn in VR.

The second point is that these drawings give evidence, whether the difference in map orientation between the different experimental conditions might come from the differences between the participants that were assigned to each condition. Here, in the training-phase, all participants walked the route in the same direction, drawing the map at the same location. Different map orientations would therefor only depend on differences between participants and not between conditions. In fact, the orientations are non-uniformly distributed with a clear preferred drawing direction (northwest=65.71%) This indicates that preferred map orientations seen in the test-phase are influenced by the factors mentioned above (4.2.2) rather than being an effect of the participants in the respective condition.

In the following discussion of the test-phase results we will distinguish between target places that were lying on the route, i.e. the *Krumme Brücke* and those that were lying off the route. In the first case, the target place was task-relevant, i.e. participants had to cross the place while following the route. This means that they either knew that the route would lead them towards the place or that they had just passed it. Therefor they would either expect to perceive a specific view (approach) or had just seen the respective view. Results show that the orientation of the sketch maps was influenced by the approach direction. That means when asked to draw the *Krumme Brücke* before they crossed it, their drawings were influenced by the view they would expect to see. When asked to draw the map after having crossed the place, drawings were influenced by the view they had just experienced.

At a first look these results seem to be only partially in agreement with the results of Röhrich et al. According to the view-graph model, drawing the map before crossing the place should induce spatial recall of the adjacent view. That means in this case participants were expected to draw the maps oriented to match the view they were expecting to see. Observed map orientations indicate that this is indeed the case. However, when drawing after having crossed the place, one might intuitively expect the maps to be oriented in the opposite direction. This is because participants standing at the second drawing location would experience the opposite view when walking towards the target place. The results show that this is not the case, map orientations are the same as those drawn before crossing the place. The reason for this lies in the connections in the view-graph and the temporal sequence of views experienced along the route. Figure 4.14 illustrates this idea. Here not only the connected views are shown, but also the direction in which participants traveled the route, i.e. if they started east or west. Thus the view experienced depends on the starting condition. For example, participants standing at drawing location B coming from the eastern starting point are experiencing view B-East. Now, when asked to draw the *Krumme Brücke*, participants will not recall the view *Krumme Brücke*-West as there is no connection to their current view. It is more likely that participants remember the view from which they arrived at their current view. This view (*Krumme Brücke*-East) again is depending on the approach direction and therefor map orientations at the second drawing location are the same as those drawn before crossing the *Krumme Brücke*.

The results are also in agreement with the findings by Basten et al. (2012). When asked to draw the sketch maps before crossing the place, participants might imagine walking towards this place. Here the situation is comparable to the mental travel induced by the priming in the previous study. Mentally traveling the route towards the target place leads m

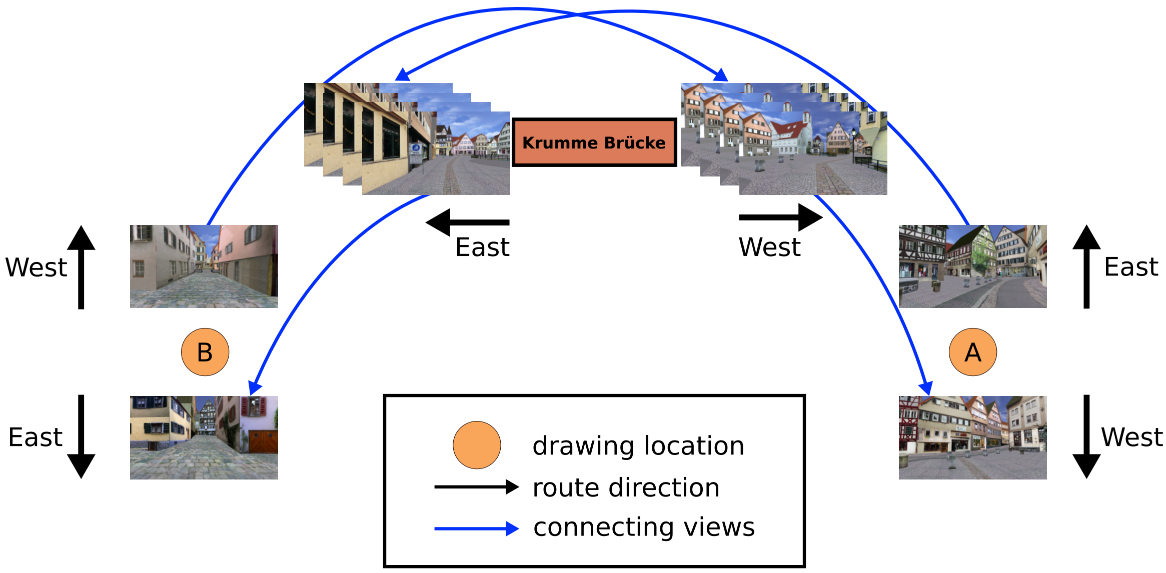


Figure 4.14: View-graph for the *Krumme Brücke* (for comparison see figure 1.3). Transitions between places are again depicted as connected views. Also shown are remembered views. E.g., participants standing at drawing location B and coming from the eastern starting point memorize the view *Krumme Brücke* $\overrightarrow{East}$

to a transfer of the respective view of that place into working memory, as confirmed by map orientations. However, when asked to draw the sketch maps after crossing the place, mental travel towards this target place would lead to sketch maps oriented in the opposite direction. The reason that the sketch maps do not show this orientation is that there is no mental travel, as participants are not instructed to imagine walking towards the place. Instead they seem to recall from episodic memory the view of the place that they have just experienced along their route.

For places that are lying off the route the situation is different. Results from the *Marktplatz* show that there is only a clear influence for the B-East condition. Here both aerial and approach direction seem to play a role. Although this place was not on the route, at the time of drawing participants were not aware of this. Instead they might have expected that following the route would lead them to the *Marktplatz*. Thus, influenced by the approach direction, they would have recalled the adjacent view connected to their current view. Therefor these results are in accordance with the view-graph model. For condition A-West and A-East there seemed to be no influence of any factor. However when looking at the histograms, it can be seen that there are two main map orientations. In both cases almost one half of the drawings are oriented west. This west direction might be explained by a canonical view, with the town hall being located in the west of the *Marktplatz*. The other half are aligned with the head direction, being roughly orthogonal to the aerial direction. This finding is in agreement with results of Meilinger et al. (2015). They found that the influence of body and location-perspective (here head and aerial direction) depended on the alignment. When both perspectives were aligned the maps were oriented along this direction. When they were misaligned orthogonally both reference frames mattered but body orientation had a greater influence. In the case of contra alignment the local perspective was more important.

The results of the *Nonnenhaus* showed an influence of the head direction for B-East and of the aerial direction for A-West and B-West. For B-East and B-West this finding is again in agreement with Meilinger et al. In one case the both directions are aligned, leading to map orientations along this direction. In the other case, the orthogonal alignment results in maps oriented along the head direction. For A-West, the alignment is almost orthogonal, but here the aerial direction seems to be more influential. However, the influence of the head direction is only weakly non-significant, also the number of participants was rather low. It is therefor reasonable not to completely neglect the influence of the head direction.

Taken together the results presented here lead to the following conclusions. First, conducting studies like (Röhrich et al., 2014) and (Basten et al., 2012) are possible using an immersive virtual reality setup. Participants were able to navigate and draw maps in a way that made the results comparable to real world studies. Second, the results are in agreement with previous studies. For task-relevant places, the map's orientation depending on the approach direction can be explained by a slightly modified view-graph model as proposed by Röhrich et al. For places that are off the route, factors like aerial and head direction are more important (Meilinger et al., 2015). Also, for these places the canonical view of the place might play a role when salient buildings or other landmarks are available.

Chapter 5

Contribution to other Work

Over the course of this thesis a lot of work was done that was not part of the experiments presented in the previous chapters. This work contributed substantially to other projects and involved programming of experimental software as well as modeling the respective environments and other required objects. In this chapter two of these projects and the contributed work will be presented exemplary.

5.1 Language Cues in Navigation

Real world environments are often divided into regions. In a study from 2003 Wiener et al. (Wiener and Mallot, 2003) could show that these regionalizations have an effect on our navigation and route planning strategies. They constructed a virtual environment that consisted of twelve places, six of them were arranged as a hexagonal ring and connected by streets. The remaining six places could be reached by dead-end roads leading away from the corners of this hexagon. At each place there was a unique landmark that distinguished it from other places. Landmarks belonged to one of three categories (cars, animals and buildings) and were arranged such that four neighboring places had landmarks of the same category. Thus the environment was divided into three distinct regions. Figure 5.1 shows an schematic overview of this arrangement.

Participants learned this environment and were then asked to find the shortest route from a given starting place to a goal place. The results show that they significantly more often chose the route which minimizes the number of region crossings. The authors therefor reasoned that "human route planning takes into account region-connectivity and is not based on place-connectivity alone".

The study which is presented in this chapter deals with the question if such hierarchical structuring of the environment can also be induced by linguistic cues (Schick et al., 2015). To examine this question a series of experiments was carried out that first aimed at repli-

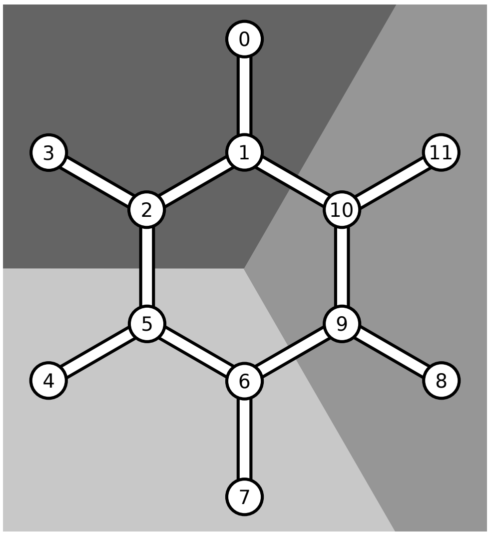


Figure 5.1: Layout of the hexagonal maze as used by Wiener (Wiener and Mallot, 2003). Six places arranged in an inner hexagonal ring, each place connected to one outer place. Regions are depicted by different shades of gray and were not visible for the participants.

cating the original results of Wiener et al. and then tested different setups of landmarks. Additionally a subsequent experiment was done to investigate the consolidation of this regionalization effect during sleep.

The problem was that the experimental software used in the study from 2003 was no longer available. Also, the framework used to program and run the experiment had been replaced over the years. Therefor the whole experiment had to be developed from scratch. New 3D models for the landmarks as well as a whole new virtual environment were required. This work, the programming of the software and the creation of all 3D models used in the experiments, was done in the course of this thesis and will be presented here in detail.

5.1.1 Software

The framework used for developing the software was again OpenSceneGraph (OSG) as in chapters 2 and 3, i.e. the code for the experimental procedure was written in C++ and OpenGL was used to render the virtual environment. The experimental procedure in general was supposed to replicate the one from Wiener et al. as accurately as possible. Therefor, when programming the experiment, a few requirements had to be considered.

The first requirement was related to the way participants navigated through the environment. In the experiments described above, they were able to move around freely within the given boundaries of the kite-shaped room. Here, in contrast, participants' movement was restricted not to leave the streets of the hexagonal maze. Therefor a navigation style was implemented that was similar to the one used in the original experiments. There were only two possible actions (rotation and translation) participants could take to navigate through the maze as follows. When standing at a place surrounded by landmarks they could turn left or right by 60°by either clicking the left or right button of the computer mouse. Therefor they could alternate between facing a landmark or a road leading away from the place. When looking down a road, participants could trigger a passive forward translation in their viewing direction by clicking the middle mouse button. If the road led to another place, then the translation ended after 100m with the participant standing at the consecutive place facing a landmark object. This was of course always the case for all places in the inner circle of the maze. Whenever participants were standing at one of the outer places facing a dead-end road, the translation ended after 10m. Participants were then automatically rotated by 180°and translated back to the place they came from, now facing a landmark object.

As described in the following section, the terrain and thus the road between two adjacent places was elevated to block participants' view towards the next place. Because of this it was necessary to adjust the way in which participants were passively translated along the road. In other experiments that were programmed using the OSG framework (Halfmann, 2010) the movement was implemented by applying translations along the *x-* and *y-axis* to the virtual camera which rendered the scene while keeping the *z-axis* (up/down direction) at zero. This method worked for the plane environments used in those previous experiments, but for traversing over the hills in the current environment the height above the ground had to be adjusted. To do this, in a first step the absolute height of the terrain at the camera's position was determined for each frame. This was done by using OSG's *osgUtil::LineSegmentIntersector*, i.e. a class that allows to compute line intersections with objects in the scene. In detail a line segment is created between a given start- and end-point and then an *osgUtil::IntersectionVisitor* is attached to the scene's camera to return all intersections. Here the start- and end-points were defined as *(x,y,10)* and *(x,y,-1)* respectively with x and y being the current camera location. The z-values were chosen to cover the whole vertical range of the environment. With no other objects being placed above the road, we assumed that the z-coordinate of the first intersection corresponded to the terrain's height. In the second step, the difference between the camera's current z-position and the terrain's height was computed and the camera was shifted up or downwards accordingly before rendering the frame.

One problem arose after a few test runs with some voluntary participants. Those participants reported that they did not have a lot of trouble learning the maze, but that they had to concentrate too much on remembering the target places they had to find. This might have impaired their navigational abilities and could therefor have an impact on the regionalization effect. Thus we needed a method that allowed participants to look up their current target places without having to memorize them. The solution was to use a second computer monitor which continuously displayed an image or text (depending on experimental condition) of the current targets. This was done by setting up a separate thread in the experimental software which was only dedicated to displaying the target places, running independent of the main program. In the experimental trials where participants had to find three different places, the thread was constantly updated so that whenever participants had reached one of the target places this place was removed from the second monitor. Also in these trials the displayed targets were constantly rotating around each other to prevent participants from being primed to a specific sequence of places. This second screen was a change to the original setup and one might argue that this display distracted too much of participants' attention. However, the primary monitor was rather large, having a size of 30 inches while the second monitor was only 14 inches in size. Thus participants' view was not constantly drawn to this display and they only looked on the second screen to their right when they felt uncertain about their current target places.

Another change that was implemented only applied to the sleep experiment and was rather an addition which made the software more easy and comfortable to use. Earlier versions of the software were simply run from the command line where the experimentalist entered parameters as, e.g., the name of the participant when prompted to do so. This approach was suitable for the standard experiment where participants came only once and did one single session, always with the same experimental setup for all participants. For the sleep experiment however, the situation was different. Here participants came twice starting either in the morning or the evening, thus having a sleep or awake phase in between two consecutive sessions. In the second session there was no exploration time given and the experiment had to be started directly into the learning phase. After a few weeks participants came for another two sessions that were run the same way, but this time with a reversed morning-evening sequence and a different virtual environment. To make the handling of all these parameters easier for the experimenter, a simple graphical user interface (GUI) was implemented. The implementation was done by using the Qt framework, i.e. an C++ class library with cross platform support for the programming of such interfaces. Within this interface, the experimentalist could enter all parameters in a comfortable way, e.g. the starting phase (exploration or learning) or the virtual environment (desert or grassland) could be chosen from a pull-down menu. After setting up all parameters those preferences could be saved and the experiment could be started. This GUI was not only easier to handle than the old command line input method, but it was also less prone to errors due to mistyping.

5.1.2 Modeling

As mentioned above, the original models of the environment and the landmarks were not available and therefor new 3D models had to be created. This was again done by using Multigen Creator©.

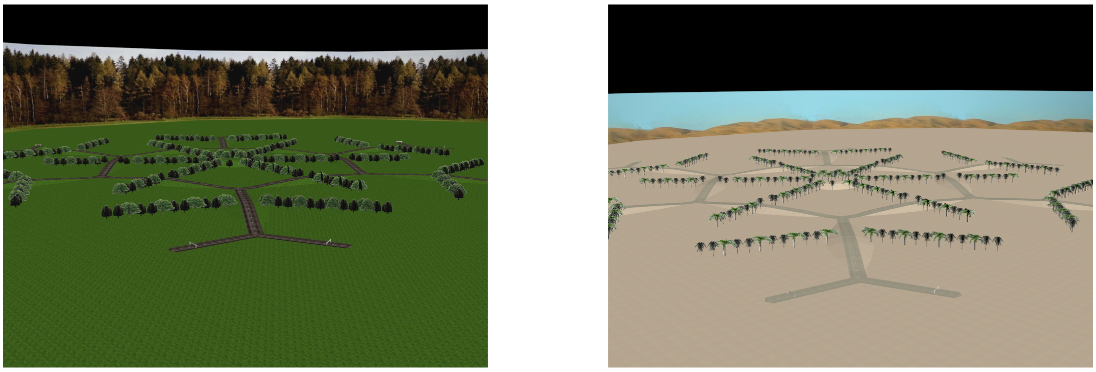
First the virtual environment was created as shown in figure 5.3a. Parts of this environment, e.g. the textures of the roads or the forest that is seen at the horizon have been used before (Halfmann, 2010; Mallot and Gillner, 2000). One constraint that had to be taken into account when designing the landscape was that when standing at a place, participants should not be able to see any other of the places from their viewpoint. In a first attempt this was simply done by implementing a fog that covered everything farther away than 25m from the observer. However, this confused participants and drastically impaired their navigational abilities. We therefor chose the same method to block the view between two adjacent places as in the original setup by placing a small hill between them. Figure 5.2 shows a direct comparison between these two methods. Additionally, to obscure the sight to the other places, rows of trees were placed between the roads. These two additions then prevented participants from using views of distant places for their navigational strategies. As mentioned in the previous section, in the sleep experiment a

(a) fog (b) hill

Figure 5.2: Example screenshot of both methods for blocking participants' view. (a) fog, starting at 25m and reaching full density at 90m. (b) elevated terrain.

second environment was needed. This environment had to be visually distinctive from the green and grassy environment but had to fulfill the same restrictions regarding the line of sight for the participants. Therefore the grass and forest background textures were replaced by sand and dune images, creating a desert-like environment (figure 5.3b). Also, the broad-leaved trees were replaced by palms which were placed the same way along the road. The layout of the environment, i.e. places, roads and hills remained identical.

For the different modifications of the experiment a different set of landmarks had to be created for each version. In the first version that aimed at replicating the results of Wiener et al., three landmark categories were chosen according to the original categories; animals, vehicles and paintings. Whereas the paintings were easily modeled by placing a



(a) grassland

(b) dessert

Figure 5.3: The virtual environments used in these experiments. (a): grassland, with forest background and broad-leaved trees, (b): desert with dune background and palms.

texture onto a canvas, vehicles and even more so the animals were too time-consuming to be modeled in a natural-looking way. We therefor used 3D models that were available for free to use and which only had to be adjusted in some details, like e.g. coat color or windshield transparency. The landmarks used in the pure language conditions were simple signs with a texture depicting the respective name of the place. However, there was one condition where language cues were combined with visual cues in the form of buildings. Here the three regions were chosen to be opera-town, church-town and court-town. Therefor for each of these towns a prototypical landmark building had to be created that represented the respective category. The other buildings were chosen to be typical buildings that could usually be found in a city but were not specific to any of the categories such as e.g. hotel, school or bakery. Examples of these landmark buildings are shown below in figure 5.4. All these buildings were created from scratch using the software mentioned above by constructing a rather low detailed 3D model and then applying high resolution textures to add a higher degree of detail. Finally, shading was applied to the models to give the final buildings a more realistic and pleasing look.

5.2 View-Specific Spatial Recall

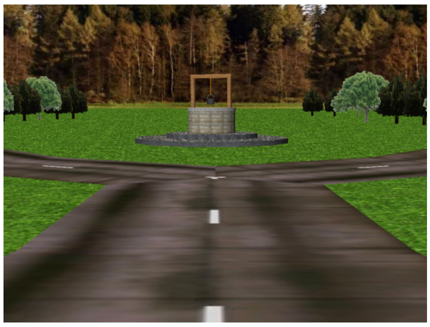
In section 1.1 it was explained how representations of places might be connected to form a cognitive map and how this stored knowledge is then transferred from longterm to working memory. Previous studies demonstrated how spatial recall is influenced by priming with imagined walking towards a place (Basten and Mallot, 2010) or by the approach direction towards a place from one's current location (Röhrich et al., 2014).

The study presented in this chapter (Haugg, 2014) investigated the effect of view-specific spatial recall on the reaction times in an image recognition task. Passers-by in the historic city of Tübingen were stopped at locations situated around the *Marktplatz* (see figure 5.5)

(a) cafe

(b) court building

(c) hotel



(d) well

Figure 5.4: Examples of the landmark buildings created for the prototype condition in Schick et al. (2015).

and asked if they wanted to participate in a short experiment. The task was then to decide whether a presented photo showed the *Markplatz* or not (yes-no-task). The *Markplatz* photos were taken from different avenues to the place. Thus there were photos that matched the view participants would perceive when walking from their current interview location to the *Markplatz* and others that did not match. Additionally non-*Markplatz* photos were added as distractor images. The idea behind this was based on the view-graph model described in section 1.1.3. That means that participants would imagine the route from their current location to the *Markplatz* and the view of the *Markplatz* which is connected to their current view should be loaded into working memory. Thus the reaction times for matching photos would be reduced in comparison to non-matching photos.

To carry out this experiment, a new setup had to be developed that allowed for maximum mobility and that was easy and intuitively to handle by the participants. This was done by using a Dell Venue 11 Pro tablet pc running windows 8. The development of this setup as well as the implementation of the software was done in the course of this thesis and will be explained in the following section.

5.2.1 Software Implementation

For the development of the experimental setup and software implementation the following requirements had to be met. First, the setup had to be mobile to allow the experimenter to go to the different interview locations in the city. Making the hardware portable was easily achieved by using a tablet pc that was lightweight and had a long battery life. The portability of the software was a bit difficult. Normally this experiment would have been programmed using the Psychtoolbox for Matlab© which already provides all necessary functions. However, the Matlab© version at that time required a constant connection to the license server of the University of Tübingen. Therefor the software would have only been usable while being within the university network. Due to the limited covering of the university's wireless network within the city of Tübingen, the usage of Matlab was not possible. Therefor the software was implemented using C++, allowing us to run a compiled version of the program independent of the network connection. As C++ does not come with its own libraries for presenting images, these functions were implemented using SDL (Simple DirectMedia Layer), a free cross-platform multimedia library.

The second requirement followed directly as a consequence of the portable setup mentioned above. As the software was running on a tablet pc with no keyboard or mouse attached, a new, touch-friendly user interface had to be developed. On the one hand, this interface should allow the experimenter to setup the experiment and enter all parameters needed. On the other hand it should be easy and intuitively usable by the participants.

First the labels for the participants were generated automatically, simply as an increasing number. To chose an interview location, the experimenter could just touch the respective marker on a city map as shown in figure 5.5. Finally, the experimenter specified

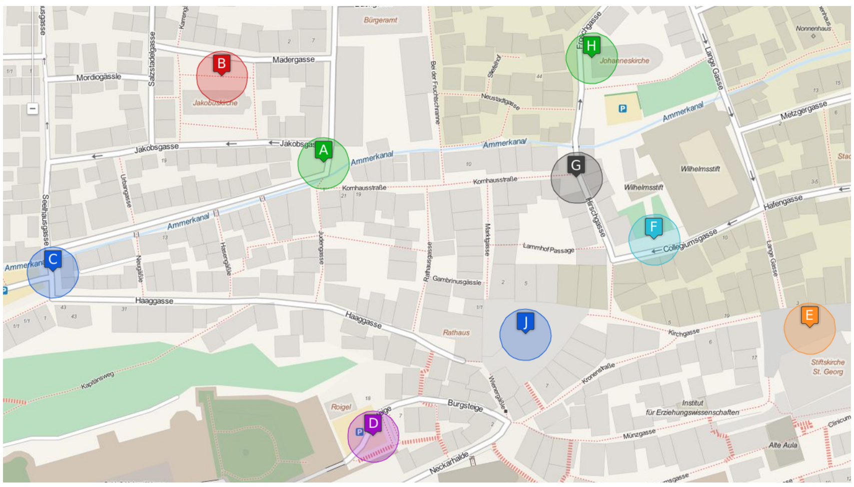
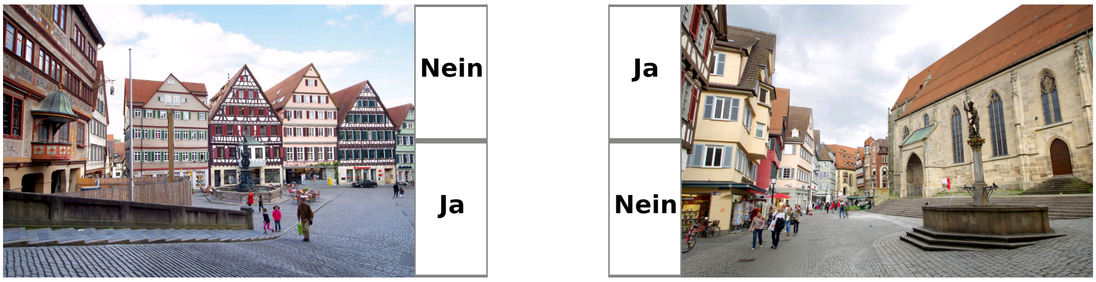


Figure 5.5: Map depicting the historic city of Tübingen. Letters A-H mark the interview locations around the *Marktplatz*. Location J was chosen to be at the *Marktplatz* itself.



(a) Marktplatz, right-handed (b) Distractor, left-handed

Figure 5.6: Interface of the experiment as presented on the tablet pc. (a) photo of the *Markplatz*, right-handed interface (b) distractor photo of the *Holzmarkt*, left-handed interface

if the participant was left or right handed, again by simply touching the respective button. Participants then held the tablet pc with their non-dominant hand while giving input with the other hand. This ensured that reaction times did not depend on being forced to use the "wrong" hand and it also made it more comfortable to use for participants. Images were than presented in a pseudo randomized, pre-defined sequence. After completing the experiment, participants were given a score (percentage of correct answers) as a little feedback about their performance.

Bibliography

Baddeley, A. (1986). *Working memory.* Oxford: Clarendon Press.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4(11):417–423.

Baddeley, A. and Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, 8:47–89.

Barry, C., Lever, C., Hayman, R., Hartley, T., Burton, S., O'Keefe, J., Jeffery, K., and Burgess, N. (2006). The boundary vector cell model of place cell firing and spatial memory. *Reviews in the neurosciences*, 17(1-2):71–97.

Basten, K. and Mallot, H. A. (2010). Simulated visual homing in desert ant natural environments: efficiency of skyline cues. *Biological Cybernetics*, 102(5):413–425.

Basten, K., Meilinger, T., and Mallot, H. A. (2012). Mental travel primes place orientation in spatial recall. In *Spatial Cognition VIII*, pages 378–385. Springer-Verlag, Berlin, Heidelberg.

Berens, P. (2009). CircStat: a Matlab toolbox for circular statistics. *Journal of Statistical Software*, 31(10):1–21.

Binder, N. (2012). *Repräsentation mentaler Karten in Abhängigkeit vom aktuellen Standort.* Bachelor thesis, University of Tübingen.

Brehm, N. (2015). *Abruf räumlicher Information aus dem Arbeits- und Langzeitgedächtnis von vertrauten Orten in virtueller Umgebung.* Bachelor thesis, University of Tübingen.

Bülthoff, H. H. and Christou, C. (2000). Using realistic virtual environments in the study of spatial encoding. In *Spatial Cognition II*, pages 317–332. Springer-Verlag, Berlin Heidelberg.

Cartwright, B. and Collett, T. (1983). Landmark learning in bees. *Journal of Comparative Physiology A*, 151(4):521–543.

Cheng, K. (1986). A purely geometric module in the rat's spatial representation. *Cognition*, 23(2):149–178.

Cheng, K. (2008). Whither geometry? Troubles of the geometric module. *Trends in Cognitive Sciences*, 12(9):355–361.

Cheung, A., Stürzl, W., Zeil, J., and Cheng, K. (2008). The information content of panoramic images II: view-based navigation in nonrectangular experimental arenas. *Journal of experimental psychology. Animal behavior processes*, 34(1):15–30.

Doeller, C. F. and Burgess, N. (2008). Distinct error-correcting and incidental learning of location relative to landmarks and boundaries. *Proceedings of the National Academy of Sciences of the United States of America*, 105(15):5909–5914.

Ekstrom, A. D., Kahana, M. J., Caplan, J. B., Fields, T. a., Isham, E. a., Newman, E. L., and Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature*, 425(6954):184–188.

Franz, M. O., Schölkopf, B., and Bülthoff, H. H. (1997). Homing by parameterized scene matching. In *Proceedings of the 4th European Conference on Artificial Life, (Eds.) P. Husbands, I. Harvey. MIT Press, Cambridge 1997*, pages 236–245.

Gillner, S., Weiß, A. M., and Mallot, H. A. (2008). Visual homing in the absence of feature-based landmark information. *Cognition*, 109(1):105–122.

Giudice, N. a., Klatzky, R. L., Bennett, C. R., and Loomis, J. M. (2012). Combining locations from working memory and long-term memory into a common spatial image. *Spatial Cognition & Computation*, 13(2):103–128.

Graham, M., Good, M., McGregor, A., and Pearce, J. M. (2006). Spatial learning based on the shape of the environment is influenced by properties of the objects forming the shape. *Journal of experimental psychology. Animal behavior processes*, 32(1):44–59.

Hafting, T., Fyhn, M., Molden, S., Moser, M., and Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806.

Halfmann, M. (2010). *Untersuchung des Einflusses von Sprache auf den Aufbau von Routenwissen*. Diploma thesis, University of Tübingen.

Hannig, M. (2012). *Entwicklung und Validierung eines spiegelstereoskopischen Virtual-Reality Aufbaus*. Diploma thesis, University of Tübingen.

Haugg, A. (2014). *Reaktionszeiten bei ansichtspezifischem räumlichen Abruf*. Bachelor thesis, University of Tübingen.

Kaplan, S. (1973). Cognitive maps in perception and thought. In *Image and Environment*, pages 63–78. Aldine.

Kollin, J. (2007). Re-engineering the Wheatstone stereoscope. *SPIE Newsroom*.

Le Conte Stevens, W. (1882). The stereoscope: it's history. *The Popular Science Monthly*, 21.

Lever, C., Burton, S., Jeewajee, A., O'Keefe, J., and Burgess, N. (2009). Boundary vector cells in the subiculum of the hippocampal formation. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 29(31):9771–7.

Loomis, J. (2013). Representing 3D space in working memory: spatial images from vision, hearing, touch, and language. In *Multisensory imagery*, pages 51–76. Springer.

Loomis, J. and Knapp, J. (2003). Visual perception of egocentric distance in real and virtual environments. *Virtual and Adaptive Environments*, (11):21–46.

Mallot, H. a. and Gillner, S. (2000). Route navigating without place recognition: What is recognised in recognition-triggered responses? *Perception*, 29(1):43–55.

Meilinger, T., Frankenstein, J., Simon, N., Bülthoff, H. H., and Bresciani, J. (2015). Not all memories are the same: situational context influences spatial recall within one's city of residency. *Psychonomic bulletin & review*, pages 1–7.

Meilinger, T., Knauff, M., and Bülthoff, H. H. (2008). Working memory in wayfinding-a dual task experiment in a virtual city. *Cognitive science*, 32(4):755–770.

MPI for Biological Cybernetics (2005). Virtual Tuebingen. *MPI for Biological Cybernetics* (online). http://virtual.tuebingen.mpg.de/index.html.

O'Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain research*, 34(1):171–175.

Parkin, S. and Macneill, J. (2014). Oculus Rift. *MIT Technology Review* (online). http://www.technologyreview.com/featuredstory/526531/oculus-rift/.

Rieser, J. J., Garing, A. E., and Young, M. F. (1994). Imagery, action, and young children's spatial orientation: it's not being there that counts, it's what one has in mind. *Child Development*, 65(5):1262–1278.

Röhrich, W. G., Hardiess, G., and Mallot, H. A. (2014). View-Based organization and interplay of spatial working and long-term memories. *PLoS ONE*, 9(11):e112793.

Schick, W., Halfmann, M., and Mallot, H. a. (2015). How to construct a linguistic landmark: language cues in the formation of hierarchical representations of space. *Cognitive Processing*.

Schölkopf, B. and Mallot, H. A. (1995). View-Based Cognitive Mapping and Path Planning. *Adaptive Behavior*, 3(3):311–348.

Siegel, A. and White, S. (1975). The development of spatial representations of large scale environments. *Advances in child development and behavior*, 10:10–55.

Slater, M., Linakis, V., Usoh, M., and Kooper, R. (1996). Immersion, presence, and performance in virtual environments: an experiment with tri-dimensional chess. *Proceedings of the 3rd ACM Symposium on Virtual Reality Software and Technology (VRST 1996), Hong Kong, China*, (JUNE):163–172.

Thompson, W. B., Willemsen, P., Gooch, A. A., Creem-Regehr, S. H., Loomis, J. M., and Beall, A. C. (2004). Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence: Teleoperators and Virtual Environments*, 13(5):560–571.

Tolman, E. C. (1948). Cognitive maps in rat and men. *The Psychological Review*, 55(4):189–208.

Vishton, J. C. and P.M. (1995). Perceiving layout and knowing distances: the integration, relative potency, and contextual use of different information about depth. *Perception of Space and Motion*, 22(5):69–117.

Wiener, J. and Mallot, H. A. (2003). 'Fine-to-Coarse' route planning and navigation in regionalized environments. *Spatial Cognition and Computation*, 3(4):331–358.

List of Figures

List of Tables

A. Appendix

A.1 Questionnaire

A.1.1 Virtual Tübingen

Fragebogen

Alter:

Geschlecht:

Wohnhaft in Tübingen für wie lange:

Einschätzung der eigenen Orientierungsfähigkeit (1= sehr schlecht, 9 = sehr gut):

1 2 3 4 5 6 7 8 9

Einschätzung der eigenen Ortskenntnisse (1= sehr schlecht, 9 = sehr gut):

1 2 3 4 5 6 7 8 9

Wie schwierig fanden Sie den Versuch (1= sehr schwer, 9 = sehr einfach):

1 2 3 4 5 6 7 8 9

Hat Ihnen der Versuch Spaß gemacht (1 = überhaupt nicht, 9 = sehr viel Spaß):

1 2 3 4 5 6 7 8 9

A.1.2 Depth Cues

Fragebogen

Geschlecht:

Alter:

Studiengang:

Kreise die Zahl die Deiner Antwort am ehesten entspricht:

Hat Dir das Experiment Spaß gemacht?

sehr wenig	1	2	3	4	5	6	7	sehr viel

Warst du motiviert?

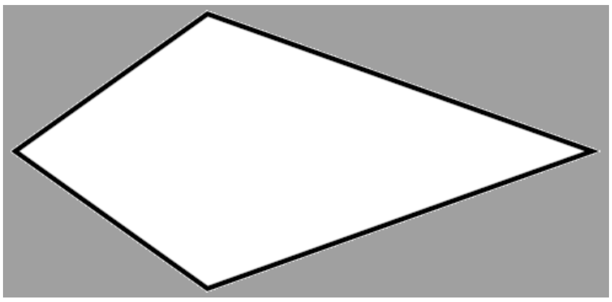
sehr wenig	1	2	3	4	5	6	7	sehr viel

Wie schwierig fandest du das Experiment?

sehr wenig	1	2	3	4	5	6	7	sehr viel

Gab es Positionen bei denen Du es einfacher fandest, sie wieder zu finden?

Kannst Du Deine Zielpunkte in den dargestellten Raum eintragen:



A.2 Figures

A.2.1 Trajectories

Baseline Experiment Stereoscope

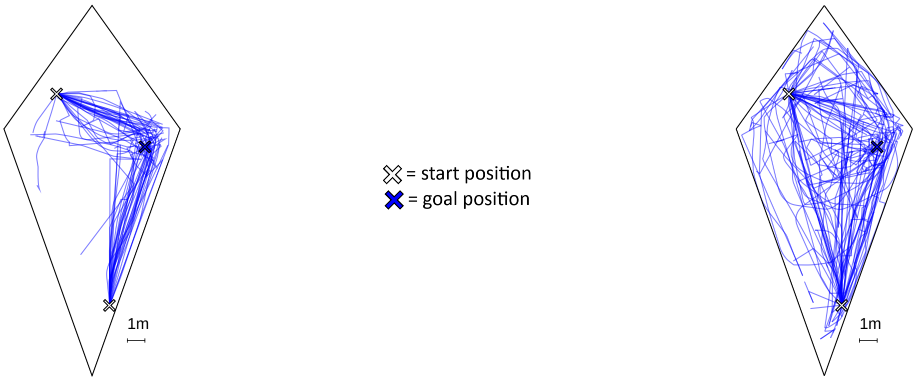


Figure A.1: Trajectories leading to goal A.

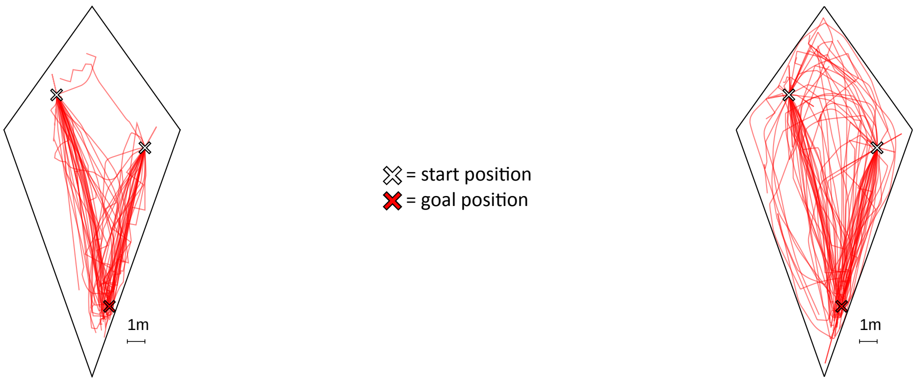


Figure A.2: Trajectories leading to goal B.

Figure A.3: Trajectories leading to goal C.

Monocular Experiment Stereoscope

Figure A.4: Trajectories leading to goal A.

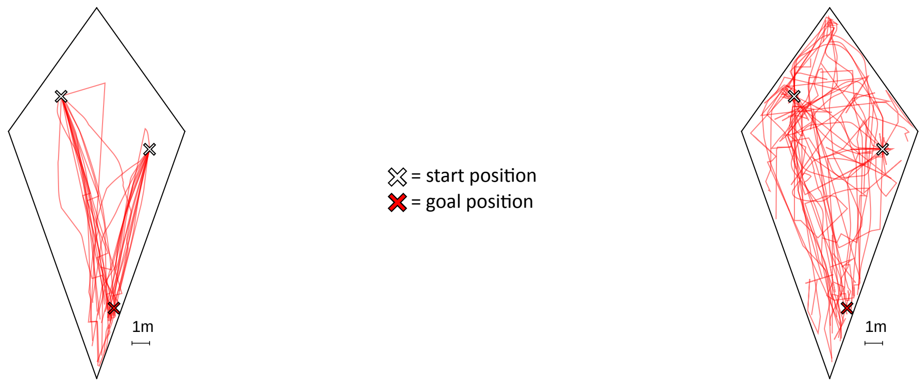


Figure A.5: Trajectories leading to goal B.

Figure A.6: Trajectories leading to goal C.

Rounded Corners Experiment Stereoscope

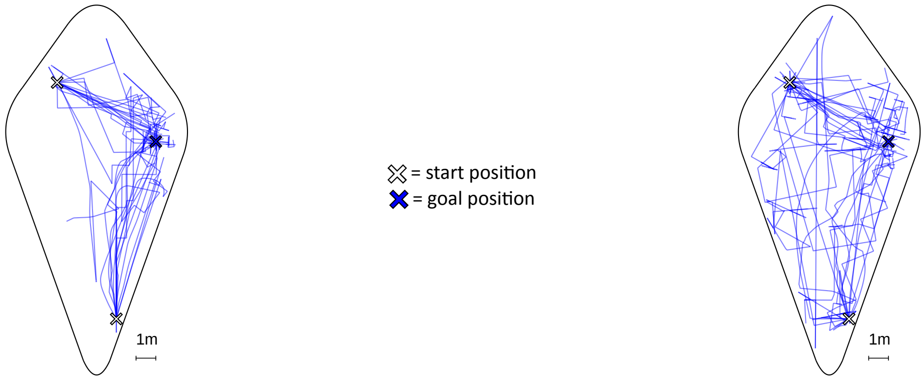


Figure A.7: Trajectories leading to goal A.

Figure A.8: Trajectories leading to goal B.

Figure A.9: Trajectories leading to goal C.

Baseline Experiment HMD

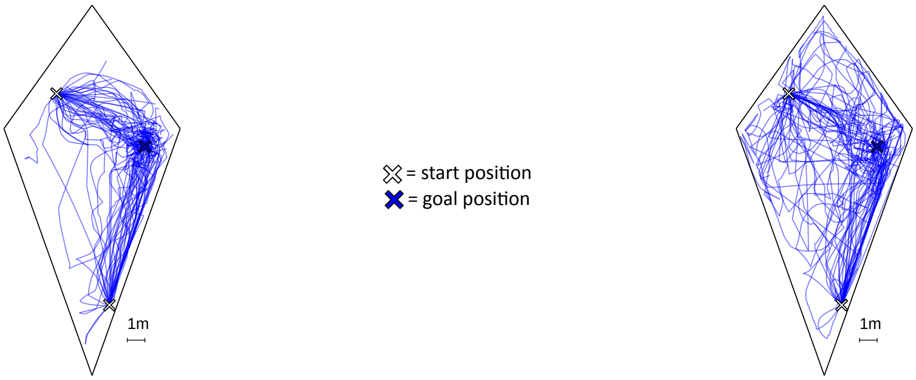


Figure A.10: Trajectories leading to goal A.

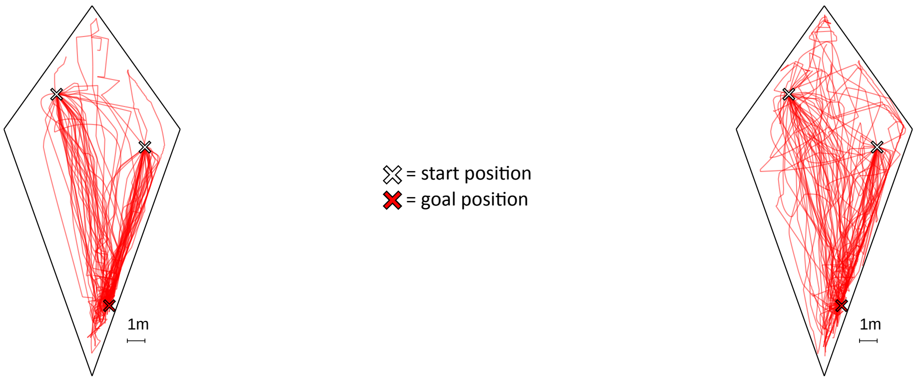


Figure A.11: Trajectories leading to goal B.

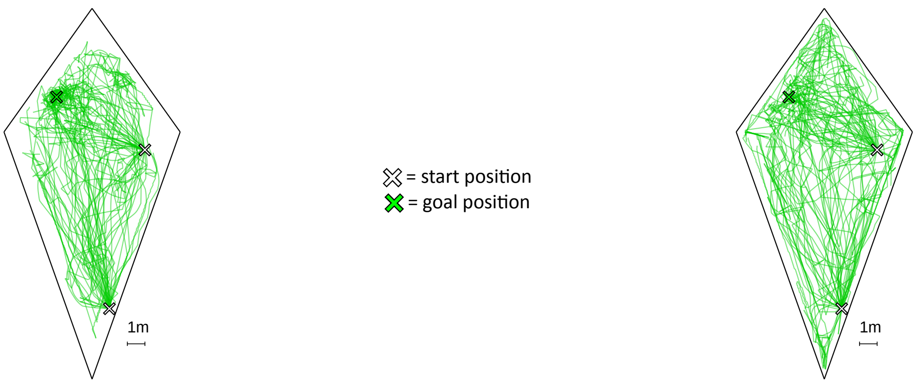


Figure A.12: Trajectories leading to goal C.

Rounded Corners Experiment HMD

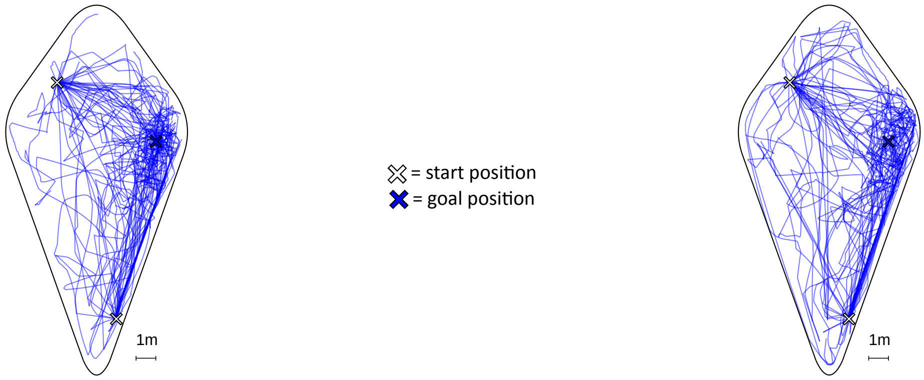


Figure A.13: Trajectories leading to goal A.

Figure A.14: Trajectories leading to goal B.

Figure A.15: Trajectories leading to goal C.