Verbal Assistance with Virtual Tactile Maps

A Multi-Modal Interface for the Non-Visual Acquisition of Spatial Knowledge

DISSERTATION
zur Erlangung des akademischen Grades
Dr. rer. nat.
an der Fakultät für Mathematik,
Informatik und Naturwissenschaften
der *Universität Hamburg*

Eingereicht beim Fach-Promotionsausschuss Informatik von

KRIS LOHMANN

aus Tönisvorst

Dezember 2012

Gutachter:

Prof. Dr. Christopher Habel
Dr. Martin Christof Kindsmüller

Hamburg, den 15.04.2013 (Tag der Disputation)

Abstract

The non-visual access to overview information about spatial environments is highly important for blind and visually impaired people. Tactile maps explored by touch can be used as substitutes for visual maps, yet, it is harder to acquire information from them than from visual maps. This thesis discusses an interface that combines situated natural-language assistance and virtual variants of tactile maps to facilitate non-visual acquisition of spatial knowledge. The haptic part of the interface, virtual tactile maps, is perceived by a map user using a haptic human-computer device. Previous empirical work on the effectiveness of combining assisting utterances and tactile maps is limited.

A set of assisting utterances for such an interface, specified based on a corpus study, is presented. Aside from information that allows to identify objects (such as: "This is the crossing between Broadway and 1st Avenue"), the set contains utterances informing the user about the local surroundings (such as: "Broadway is leftwards restricted by the map frame and forms a dead end to the right. Above Broadway, there is the town hall. ...").

An evaluation of the effectiveness of assisting utterances to facilitate non-visual spatial knowledge acquisition from virtual tactile maps is discussed. Three controlled experiments were conducted with blindfolded sighted and a mixed group of blind and visually impaired participants, who are the potential user group.

The empirical results show that, in general, assisting utterances facilitate the non-visual acquisition of knowledge from virtual tactile maps. This shows the potential of the combination of modalities to communicate information non-visually. Blind and visually impaired people acquire more knowledge from virtual tactile maps receiving assisting utterances than when they only receive information about the names of objects verbally. The empirical results show that blindfolded sighted people perform differently than blind and visually impaired people.

A second goal of this thesis addresses the technical possibility of a system that generates assisting utterances as proposed. Such assisting utterances for explorations of virtual tactile maps are situated; that is, their content depends on the context of their articulation. The generation of situated assisting utterances is a technically challenging task, to which a solution is discussed.

With respect to the technical plausibility of the assistance-generation system, an architecture for such a system is presented. This thesis is centered on the language-generation component of the system. Depending on the component detecting hand movements, the timing of the input to the generation component can vary. It is shown that a prototype of the generation component is robust against differently timed input. Furthermore, it is shown that the component works when integrated in a prototype of the complete system.

To sum up, the results show the potential and the practicability of using natural language in combination with virtual haptics to present spatial information non-visually. The results imply that blind and visually impaired people are able to acquire integrated spatial knowledge. Furthermore, the results point towards differences in the integration of spatial language and haptic information between visually impaired and sighted people.

Zusammenfassung

Der nicht-visuelle Zugang zu Informationen über räumliche Umgebungen ist äußerst wichtig für blinde und sehbehinderte Menschen. Erfühlbare taktile Karten ermöglichen den nicht-visuellen Zugang zu Informationen, sind aber schwerer zu erfassen als visuelle Karten. Diese Dissertation beschäftigt sich mit einem Interface, dass sprachliche Assistenz und virtuelle taktilen Karten kombiniert und so den Erwerb von Überblickswissen für blinde und sehbehinderte Menschen vereinfacht. Der haptische Teil des Interfaces, virtuelle taktile Karten, kann mit einer haptischen Mensch-Computer-Schnittstelle wahrgenommen werden. Die bisherigen empirischen Ergebnisse, die sich mit der Kombination von räumlicher Sprache und taktilen Karten auseinandersetzen, sind begrenzt.

Ein Set assistierender Äußerungen, das auf der Basis einer Korpus-Studie erstellt wurde, wird diskutiert. Neben Informationen, die es erlauben Objekte zu identifizieren, vermitteln die Äußerungen zum Beispiel Informationen über die lokale Umgebung des Kartenobjekts.

Eine Evaluation der Effektivität assistierender Äußerungen zur Vereinfachung des nicht-visuellen Wissenserwerbs wird vorgestellt. Dazu wurden drei Experimente mit sehenden Teilnehmern mit verbundenen Augen und mit blinden und sehbehinderten Teilnehmern durchgeführt.

Die empirischen Ergebnisse zeigen, dass sprachliche Assistenz den nicht-visuellen Wissenserwerb im Allgemeinen unterstützt. Dies zeigt das Potential der Kombination der Modalitäten, Informationen nicht-visuell zu vermitteln. Blinde und sehbehinderte Menschen bilden die potentielle Benutzergruppe. Sie erwerben mehr Wissen von virtuellen taktilen Karten, wenn sie entsprechende assistierende Äußerungen erhalten, als wenn sie nur Informationen über die Namen der Kartenobjekte erhalten. Die letztere Verwendung von sprachlichen Äußerungen ist vergleichbar zu früheren audio-taktilen Karten-Systemen. Die Resultate zeigen auch, dass sehende Menschen und blinde und sehbehinderte Menschen sich im Wissenserwerb unterscheiden.

Ein zweites Forschungs-Ziel befasst sich mit der technischen Machbarkeit eines Systems, das automatisch sprachliche Assistenz generiert. Assistierende Äußerungen sind situiert, das heißt, ihr Inhalt hängt vom Kontext der Artikulation ab. Die Generierung situierter assistierender Äußerungen ist eine technisch anspruchsvolle Aufgabe, für die eine Lösung beschrieben wird.

Es wird eine Architektur für ein System, das sprachliche Assistenz generiert, präsentiert. Diese Arbeit fokussiert auf die Sprach-Generierungs-Komponente eines solchen Systems. Es wird gezeigt, dass sich der Prototyp der Komponente robust gegenüber verschiedenen Input-Typen verhält und dass er im Gesamtsystem funktioniert.

Zusammengefasst zeigen die Resultate das Potential und die Umsetzbarkeit der Nutzung von Sprache in Kombination mit haptischen Schnittstellen zur nicht-visuellen Darstellung von räumlicher Information. Die Ergebnisse zeigen die Fähigkeit von sehbehinderten Menschen, integriertes räumliches Wissen zu erwerben. Des Weiteren deuten die Ergebnisse darauf hin, dass sehbehinderte Menschen Sprache und haptische Darstellungen anders integrieren als sehende Menschen.

Acknowledgments

I thank my 'Doktorvater' Christopher Habel for intellectual support and for the excellent supervision of this thesis. Likewise, Carola Eschenbach deserves greatest thanks for her detailed comments and valuable and helpful feedback.

I am grateful to Martin Christof Kindsmüller for accepting to review this thesis and for his valuable and constructive feedback.

The work reported has been financially supported by the German Science Foundation (DFG) in the international research training group 'Cross-modal Interaction in Natural and Artificial Cognitive Systems' (CINACS), IRTG 1247, which has made this interdisciplinary thesis possible. I thank Jianwei Zhang, who is the the spokesperson of CINACS.

Special thanks go to Patrick Stahl for being my intellectual sparring partner since our undergraduate studies. I thank my colleagues in CINACS and in the Knowledge and Language Processing Group (WSV). We spent some time that I will not forget. In particular, I thank Stephanie Badde and Christian Floß for taking time for discussion of the details of my empirical data collection. In the same direction, I wish to express my thankfulness towards Benny Haarhaus (not part of CINACS). I thank Felix Lindner, Junlei Yu, and Patrick McCrae for valuable remarks on parts of this thesis. Timo Baumann, Matthias Kerzel, and Junlei Yu (again) deserve thanks for our successful collaboration. Ole Eichhorn and Philipp Schlesinger have supported this work as CINACS research students.

There are more people who supported me with valuable discussions; in particular, I would like to mention Sven Degenhardt, Mark May, Nick Giudice, and Thora Tenbrink.

The people from Louis-Braille-Center in Hamburg, in particular, Ulli Staniullo and Carsten Albrecht deserve thanks for being open for discussion and allowing me to use their rooms as laboratory for a part of my experiments. Tatjana 'Lu' Tetsis and Hildegard Westermann deserve greatest thanks for all the help with administrative tasks.

I thank Mirjam Krapohl for supporting me, giving me confidence, being so patient, and in general, for being there for me. My family—that is, my parents Christine and Michael and my sisters Luzie and Mieke Lohmann—has greatly supported me without hesitation during my education, including the work that has led to this thesis.

Contents

Chapter 1

Introduction

1.1 Motivation

Humans frequently rely on maps and their electronic counterparts, (online) maps such as OpenStreetMaps[1] and Google Maps[2]—for example, to find out where a university they visit for a talk is. As survey representations of spatial information that show an overview over large environments, maps enable and facilitate spatial decision making. Maps are important external representations, because they represent environments that are too large to be perceived from the limited view point of a human in the environment. If humans have no maps at hand, they have to move through large environments to gain survey knowledge about them (unless in rare occasions, when seen from a bird's-eye perspective, e.g., from a plane).

Usually, maps and their electronic counterparts are perceived using sight. Because visual maps and vision-based geographic information systems are not accessible to visually impaired people[3], tactile maps are used as substituting means. In traditional tactile maps, raised surfaces allow to explore them by touch. One sensory modality (visual representation), which can represent space, is replaced by another modality (representation for touch), which also can represent space. Yet, vision has not without reason been called the "spatial sense par excellence" (Foulke, 1982, p. 61): It is the primary sensory modality of spatial cognition (Eimer, 2004; Cattaneo et al., 2008; Cattaneo & Vecchi, 2011). Vision is highly parallel with respect to the perception of spatial information. Although vision relies on eye movements, with respect to maps, humans can access the information in the visual field nearly simultaneously (Cattaneo & Vecchi, 2011). Vision allows for both a large amount of simultaneously available information and for high selectivity (Foulke, 1982). Because physical contact is not needed, it is possible to perceive objects farther away than reaching distance (such as large distant buildings).

In contrast to vision, the haptic sense has a more local character: Distant objects cannot be perceived haptically. In addition, the haptic sense has a more sequential character than vision, at least with respect to highly structured representations of space such as tactile maps (Loomis, Klatzky, & Lederman, 1991; Ungar, Blades, & Spencer, 1995; Röder,

[1]http://www.openstreetmaps.org

[2]http://maps.google.com

[3]In the literature, it is sometimes distinguished between blind and visually impaired people (see de Almeida (Vasconcellos) & Tsuji, 2005, for a discussion). I use the term 'visually impaired people' to refer to both, totally blind people and people with some remaining visual perception (such as dark/light).

Rösler, Grunwald, & Beyer, 2001; Cattaneo et al., 2008; Loomis, Klatzky, & Giudice, 2012, but c.f. Morash, Connell Pensky, Alfaro, & McKerracher, 2012, for a discussion that not all kinds of haptic perception are sequential). Due to the sequentiality of haptic perception, tactile maps have to be explored part by part. The sequentially acquired information has to be integrated to form coherent spatial knowledge. Additionally, the possible information density on tactile maps is more sparse than on visual maps.

Reading a tactile map can be more time consuming and complicated than visual map reading (e.g., Lawrence, 2011). Furthermore, on tactile-only, uni-modal maps problems to represent the names of the map objects occur because tactile Braille writing is too inflexible to be used on maps (Jacobson, 1996).

Suitable means to acquire spatial knowledge are crucial for visually impaired people. As Golledge (1993, p. 71) points out: “Whether congenitally blind, adventitiously blind, otherwise visually impaired, or sighted, our general quality of life depends to a large degree on our ability to make spatial decision[s].”

The drawbacks of tactile maps in terms of speed and accuracy are an important topic with respect to the ability to make spatial decisions, which, as discussed, are enabled and facilitated by representations of environments. In this thesis, it is suggested to utilize natural language in the form of assisting utterances to reduce the drawbacks of tactile maps. The utterances suggested include information about the names of objects and, thus, address the problem of how names of objects can be represented in tactile maps (giving names via speech corresponds to the current state of the art in the development of audio-tactile maps, see Chapter 2).

Not all objects on a map have proper names. Assisting utterances can be used to inform the map user about the identity of objects without proper names, as well. For humans, it is common to produce and understand utterances referring to objects without a proper name, such as intersections: ‘This is the crossing between De Grassi Street and 1st Avenue.’ Constructing such expressions that allow for identification of an object without a proper name is potentially helpful to acquire knowledge from maps. Because deictic expressions that identify an object can establish the identity of objects that are currently haptically explored, these kinds of expressions can potentially reduce the problem of getting “lost in haptic space” (Colwell, Petrie, Kornbrot, Hardwick, & Furner, 1998a, p. 97) in virtual haptic environments. With respect to tactile maps, getting lost means that the users do not know which map object or which part of the map they currently explore. The identification of the object currently explored potentially reduces this problem.

Moreover, as an elaborate representational system, natural language can communicate information that is not perceived haptically at the time of utterance. Assisting utterances can be used to convey information about spatial relations of map objects (such as that a landmark is to the left of another landmark) and of configurational information (e.g., listing all intersections a street has) when the map user is not currently exploring the respective objects linguistically referred to. Therefore, in addition to informing the map user about the names of objects, assisting utterances can possibly reduce the drawbacks resulting from the sequentiality of the tactile-map reading process and the sparseness of information on tactile maps.

The use of assisting utterances to facilitate explorations of tactile maps has not been investigated systematically, neither with respect to appropriate techniques for the automatic generation of situated assisting utterances nor with respect to empirical evaluation of their effectiveness. In general, work on the combination of two representational systems (such as maps and language) is sparse (see Chapter 2). The gap in earlier research includes appropriate evaluation procedures for interfaces used to acquire survey knowledge non-visually.

The goal of the research reported is to evaluate the usefulness and feasibility of verbal assistance for virtual variants of tactile maps—called 'virtual tactile maps' in the following. I discuss a set of assisting utterances for explorations of such maps based on how humans solve the task to assist a map explorer. I report on an evaluation of the usefulness of assisting utterances to facilitate spatial knowledge acquisition of virtual tactile maps including a discussion of experimental methods. Furthermore, I discuss a technical solution for a component that automatically solves the task to generate assisting utterances for explorations of virtual tactile maps.

1.2 Virtual Tactile Maps and Assisting Utterances

1.2.1 Verbally Assisting Virtual-Environment Tactile Maps (VAVETaMs)

Verbally Assisting Virtual(-Environment) Tactile Maps (VAVETaMs) are proposed as an interface for non-visual access to information about environments. VAVETaMs combine two ways to represent space: virtual tactile maps and natural language in the form of assisting utterances. Virtual tactile maps present spatial information similar to a visual map. As in a visual map, the environment is projected on a two-dimensional plane. While in a visual map colors are used to display map objects, such as the streets and buildings, in (virtual) tactile maps[4], map objects have to be displayed for touch—for instance, by raising or indenting the areas that represent them.

To enable the perception of virtual tactile maps, a commercially available device that allows for haptic perception of virtual objects is used. This device is called 'Phantom Omni'[5]. It is discussed in detail in Section 2.3.1. The virtual haptic objects presented are based on three-dimensional models such as used in visual virtual-reality environments or computer games. In the case of VAVETaMs, in which the device is used to present spatial information non-visually, the virtual objects are virtual tactile maps[6].

[4]Throughout the text, I use the term '(virtual) tactile maps' when the arguments I present generalize to both, traditional (physical) tactile maps and their virtual counterparts, virtual tactile maps. If the argument does not generalize to tactile maps, I use the term 'virtual tactile map' without parentheses around 'virtual'.

[5]http://www.sensable.com

[6]Haptic perception with the Phantom device is mostly based on kinesthetic feedback. However, it is not only based on kinesthetics, but also on some cutaneous feedback (I thank Nick Giudice for this remark). Hence, the term 'virtual tactile maps' is justified and used in this text. The term 'tactile map' is an established term for external representations of space explored by touch.

Large-scale Braille displays are used in research (e.g. Zeng & Weber, 2012), but are too expensive for individual use[7]. Furthermore, the maximal spatial resolution of such displays is much lower than that of the Phantom device. The haptic interaction with Phantom devices is comparably sparse: It is similar to exploring an object with a pen (O'Malley & Goldfarb, 2002; O'Malley & Uppermann, 2006). Yet, in comparison to gaming joysticks successfully used in earlier research (Lahav & Mioduser, 2008a, 2008b), Phantom devices provide a realistic spatial experience. Phantom devices have successfully been used to present maps for visually impaired people (Kostopoulos, Moustakas, Tzovaras, & Nikolakis, 2007; De Felice, Renna, Attolico, & Distante, 2007). They are commercially available haptic devices that allow the construction of spatial representations in an affordable price-range for individual users (aside from tactile printing systems, which impose the need to produce a physical copy of the map area of interest and, therefore, are inflexible, costly, and time-consuming to use).

As Kaklanis, Votis, Moschonas, and Tzovaras (2011) discuss, Phantom-based virtual tactile maps can be generated from online available geographic information (e.g., from OpenStreetMaps). Thus, virtual tactile maps overcome the problem of availability and costly production of tactile maps.

The acronym 'VAVETaMs' refers to the external representations, not to the system with which they are realized (the VAVETaM system is discussed in the next section). The users of VAVETaMs receive assisting utterances when they explore the virtual tactile map. For example, when they explore a street called 'Broadway', they receive: 'This is Broadway. It runs parallel to Dufferin Avenue. Above Broadway, there is the town hall and the university. Broadway is upwards restricted by the map frame and downwards forms a dead end. Broadway has intersections with 3$^{\text{rd}}$ and 4$^{\text{th}}$ Street.' When they explore landmarks such as buildings, the users receive similar assisting utterances.

The goal of using VAVETaMs is to acquire an overview of the area presented. Such knowledge is called 'survey knowledge'. Survey knowledge is needed for robust navigation and autonomous (pre-journey) planning of a route. As survey representations, maps are especially adequate to acquire survey knowledge. Consequently, the effective non-visual acquisition of this kind of spatial knowledge is the primary goal of the research reported. The problem of non-visual route guidance (i.e., providing online-information during navigation) is a much better investigated area (Giudice & Legge, 2008) and commercial non-visual guidance systems based on GPS are available[8].

The maps investigated represent urban environments. Important types of objects in such environments are streets and other locomotion-enabling structures and landmarks, such as trees or buildings. The users explore VAVETaMs freely; that is, they are not restricted with respect to the order in which they explore parts of the maps and objects on the maps. Accordingly, map users can focus on the parts of the map they are most interested in or that are most relevant for them.

[7] http://www.hyperbraille.de/file_download/1/flyer_hyperbraille.pdf

[8] http://www.senderogroup.com

1.2.2 The VAVETaM System

To present the haptic part of VAVETaMs and to generate assisting utterances, we[9] suggested an automated system called 'VAVETaM system' (Lohmann, Kerzel, & Habel, 2010; Habel, Kerzel, & Lohmann, 2010). The usage scenario is that a map user explores VAVETaMs to acquire survey knowledge of an urban area. Users explore the spatial representation to acquire knowledge about the environment that is shown. In this respect, the usage scenario of the VAVETaM system is comparable to the use of a traditional map or online maps. The virtual tactile maps and the assisting utterances are presented and generated by the VAVETaM system. The map user receives assisting utterances that are generated on the basis of his or her explorative movements.

My research goal with respect to the VAVETaM system is to show that the generation of assisting utterances for free, unguided explorations of tactile maps is technically possible; that is, to show the feasibility of such a project. Therefore, a component solving the assistance-generation task was developed and implemented in a prototype. The assisting utterances are situated; that is, their generation takes the current context into account. The situated generation of assisting utterances is a complex task that demands for different interacting components. On a coarse-grained level, two important tasks have to be solved: Analyzing the maps user's explorative hand movements (Kerzel & Habel, 2011, 2012) and generating situated assisting utterances (Lohmann, Eschenbach, & Habel, 2011; Lohmann, Eichhorn, & Baumann, 2012). My research concerns the second task and the interface between the components.

1.3 Disciplines and Methods

Human-computer interfaces should not only be technically feasible and adequate, they should also be adequate with respect to the perceptual and cognitive abilities of potential users (e.g., Loomis et al., 2012)—that is, they should be understandable and effective. Interdisciplinary work allows to combine the task to show that an interface is effective and useful and the task to show that it is technically possible.

The work reported is based on and influenced by methods from different scientific areas, most importantly, human computer interaction, cognitive science, spatial cognition, and natural language generation (a subfield of computational linguistics). These areas are themselves interdisciplinary and based on methods of scientific disciplines such as psychology and computer science. Human computer interaction is concerned with the design, development, and evaluation of interfaces to control a computer. Human computer interaction is central to the research reported because VAVETaMs are interfaces to geographic information stored on computers.

[9] I am responsible for the work reported. Yet, this work was discussed with my supervisors and colleagues and parts of it were collaborative (especially, the implementation of the prototype system reported in Section 4.8 is a combination of an event detection component implemented by Matthias Kerzel and the work reported in this thesis). Therefore, I use the editorial 'we' in cases where I describe work that has been done ('We tested ...'). However, I wish to make clear that all possible mistakes and flaws are to be attributed to my person and not to the people I collaborated with. In cases where I refer to the text ('I discuss the term in Section ...'), I use the singular 'I'.

Cognitive science is concerned with the study of the human mind and intelligence. Cognitive science investigates the processes that underly human thinking and reasoning. Research in cognitive science combines different methods such as neuropsychological, psychological, linguistic, and philosophical methods.

The goal of spatial cognition is to investigate how humans (and other animals or artificial systems) acquire spatial knowledge, how they structure spatial knowledge, and which processes are involved. VAVETaMs are used to acquire spatial knowledge. With respect to this specific domain, spatial cognition is of important interest, especially, because adequate methods for testing spatial knowledge with respect to the interaction scenario have to be developed. Therefore, in particular, the evaluation procedures (i.e., the experimental design and the testing methods) and the theoretical background on humans' spatial knowledge are based on research in the area of spatial cognition. In addition, the empirical research reported relies on user studies in Wizard-of-Oz-like settings (in which an experimenter-controlled system is tested allowing testing prior to implementation) and with prototypes—methods used in human computer interaction.

With respect to the question how a situated natural-language generation system can be realized, the work reported has its roots in the research area of natural language generation. Natural language generation is concerned with the development of artificial systems that have the capability to generate natural language. Additionally, some ideas are based on psycholinguistic research, which is concerned with how humans are able to understand and produce natural language. Especially the architecture of the generation part of the VAVETaM system and the interfaces between its subcomponents take findings both from psycholinguistics and natural language generation into account.

A crucial part of a language-generation system is how semantic knowledge is represented. In the language-generation part of the VAVETaM system, the representation of the semantics of assisting utterances is based on the referential-nets formalism—a formalism developed in the areas of artificial intelligence and cognitive science (see Habel, 1986, a discussion of the relation of these scientific areas can also be found there).

Some of the research questions investigated are related to earlier work in cognitive science and, again, psycholinguistics; for example, the idea of reference systems used in human processing of spatial information is rooted in these disciplines. In addition, cognitive science and philosophy of media are important sources for my definition of different forms of representations and multi-modality in representations.

1.4 Overview of the Structure

My research addresses the helpfulness and feasibility of augmenting virtual tactile maps with assisting utterances. I address different research questions, mainly in the following areas: the development of helpful verbal information, the evaluation of the effectiveness of VAVETaMs, and the appropriateness of the technical solution proposed for the VAVETaM system. Specifically, the research questions I address are:

1. What should be said and when should it be said to support non-visual knowledge acquisition of virtual tactile maps?

2. Which methods should be applied to empirically evaluate the approach?
3. Do assisting utterances facilitate non-visual knowledge acquisition of virtual tactile maps?
4. Can a technical solution be used to generate assisting utterances?

This thesis is structured as follows: After a review of the literature and the introduction of important terms, it is investigated how humans solve the task to verbally assist a user of a virtual tactile map. I discuss a set of potentially helpful assisting utterances created on the basis of the data collected in a corpus study. Then, this set of assisting utterances is evaluated in controlled experiments. In the last step, I show the technical possibility of generating assisting utterances as suggested by developing a prototype solving this task.

In more detail, Chapter 2 discusses the nature of spatial representations. I discuss definitions for the term 'representation' and analyze different classes of representations; in particular, I define 'representational multi-modality'. Both (tactile) maps and natural language can serve as sources for spatial knowledge. I begin the discussion of relevant representations of space with tactile maps and the virtual tactile maps that are the haptic part of VAVETaMs. Then, earlier approaches towards audio-tactile maps are reviewed. The literature review indicates that up to now, no systematic evaluation of the use of assisting utterances in combination with (virtual or traditional) tactile maps has been made. Furthermore, evaluations of audio-tactile maps in controlled experiments are sparse. I also review the relevant literature on how sighted and visually impaired humans structure spatial knowledge. Then, I review literature on how humans, in particular visually impaired individuals, acquire spatial knowledge from different sources.

Chapter 3 is concerned with the research questions 1 to 3. I present a set of potentially helpful assisting utterances based on the corpus study mentioned. Three controlled experiments, which evaluate whether this set facilitates knowledge acquisition, are discussed. The experiments were made before implementing the VAVETaM system (the assisting utterances were controlled by the experimenter in a Wizard-of-Oz-like setting).

Chapter 4 discusses how the generation of natural language in the form of assisting utterances can be realized, as addressed by research question 4. As mentioned, the user explores VAVETaMs freely. Consequently, the utterance generation has to be situated with respect to what the map user currently explores and which information has been given already. The generation component is discussed and I illustrate how it works. As reported in Lohmann, Kerzel, and Habel (2012), we integrated an event detection component and the natural-language generation component discussed in this thesis to a prototype of the VAVETaM system. This prototype was evaluated in a user study. The relevant results of the user study are discussed.

I conclude with Chapter 5, which summarizes the findings and discusses some open questions.

Chapter 2

Maps and Spatial Knowledge

2.1 Introduction

Visual maps, (virtual) tactile maps, and descriptions of space are representations of spatial environments. They have very different properties: While maps are themselves spatial, information in descriptions is organized sequentially. Maps and descriptions are combined in VAVETaMs, which are used to acquire spatial knowledge non-visually. Humans structure their spatial knowledge mentally (they go beyond stimulus-response learning) and organize it in a representation of the spatial environment. As this brief discussion shows, the concept of representation is central for the investigation of VAVETaMs.

The term 'representation' has various meanings, especially in different scientific contexts. In cognitive science, it is mainly used to refer to mental structures that preserve information about objects or events in the world (McNamara, 1999; Pitt, 2008). In computer science, 'representation' refers to the way information is stored and structured as data by computing machinery (Poole & Macworth, 2010). In philosophy (and philosophically oriented linguistics), representations are studied in a broader context. For example, the field of semiotics is concerned with how something can refer to something else in general (Peirce, 1960, 2.228).

I use the term as defined by Palmer (1978, p. 262): "A representation is [...] something that stands for something else." That representations stand for something else is commonly accepted. It is the starting point of McNamara's (1999) discussion of representations, as well. Furthermore, it is very similar to Peirce's definition of signs[1] (yet, this definition includes that a sign must be interpreted by someone to function as a sign; Peirce, 1960, 2.228). According to Palmer, that for which a representation stands is the represented world. The representation and the represented world are the parts of a *representational configuration*.[2] The representation does not have to stand for all aspects (i.e., in other words, properties) of the represented world. For something to be a representation, it is sufficient that it represents some aspects of the represented world.

[1] In Peirce's work, the term 'sign' is used almost interchangeably with the term 'representation' (in the earlier works of Peirce signs are frequently called 'representations', Atkin, 2010).

[2] In Palmer's text, it remains ambiguous whether the term 'representation' refers to both, the representing world (called 'representation' here) and the represented world, or whether it just refers to the representing world. In order to make a coherent use of the terms, I use 'representation' only to refer to the representing world. To refer to both, the represented world and the representing world, I use the term 'representational configuration'.

For example, a typical road map stands for the course and length of the streets represented (here, the streets are the represented world). Such a road map does not include information about altitude and the resulting slope of the streets, even though these are aspects of the represented world—that is, of the real-world streets (which the representation stands for). Still, the road map is a representation of the real-world streets (see also Habel, 1996, 2003b).

The definition discussed shows the extension (i.e., the things which it applies to or for which it stands) of the term 'representation': The term refers to everything that stands for something else, including mental representations that humans generate from perceptions and use in every-day reasoning. As well, the term is used to refer to physical[3] things that stand for something else. The former, humans' mental representations, are called 'internal representations' and the latter, physical representations, are called 'external representations'.

Humans use diverse external representations of space to acquire spatial knowledge, such as maps and verbal descriptions (Liben, 1981, 1991; Habel, 2003a; Newcombe & Huttenlocher, 2003). External representations of space play a major role in solving tasks such as planning the way to a conference venue in a novel environment (the potential of external representations to facilitate problem solving is discussed in Scaife & Rogers, 1996).

This chapter is structured as follows. To introduce and define the terminology, it starts with introducing analog, propositional, and symbolic representations (Section 2.2.1). Representations are often combinations of analog and symbolic representations. Such representations are called 'hybrid representations' (Section 2.2.2). External representations can be perceived via different sensory modalities such as audition and sight. Representations that combine different sensory modalities are called 'sensory multi-modal representations'. However, external representations can also be multi-modal with respect to the representational modalities (natural language, maps, ...) they are based on. Such constellations are discussed in Section 2.2.3.

Spatial environments can be non-visually represented by different external representations, such as representations based on natural language or tactile maps. For this thesis, it is important that research has shown that tactile maps can effectively be used by visually impaired people as substituting means for visual maps, aside from the drawbacks discussed (Section 2.3.1).

Spatial information is encoded in natural language on the basis of reference systems. Equally, maps are organized on the basis of reference systems (e.g., the cardinal directions establish a frame of reference). In a representationally multi-modal setting including a (virtual tactile) map and assisting utterances, the assisting utterances have to adopt a reference system that is appropriate (Section 2.3.2).

The state of the art in audio-tactile maps developed for non-visual knowledge acquisition is reviewed in Section 2.3.3. Essentially, the successful use of external representations of space results in the acquisition of spatial knowledge—that is, an internal representation of the environment. Humans' can acquire integrated spatial knowledge in the form

[3]The term 'physical' is used in a broad sense here. For example, it includes non-persistent pictures on computer screens that are created on the basis of data.

of survey knowledge. The corresponding internal representations of spatial knowledge are called 'spatial mental model' in this thesis (Sections 2.4.1 and 2.4.2). In Section 2.4.3, I discuss the content of spatial mental models of cities. The current state of research indicates that blind and visually impaired people are able to build up spatial mental models (Section 2.4.4), which is important, as the goal of VAVETaMs is to communicate survey knoweldge. Research has shown that both spatial natural language and tactile maps can be used effectively by visually impaired people to acquire spatial knowledge. This indicates that both representational modalities used in VAVETaMs can be used unimodal. Yet, no systematic evaluation of combinations of these two modalities to a representationally multi-modal interface to spatial information has been made yet (Section 2.5.2).

2.2 Properties of Representations

2.2.1 Classes of Representations

Figure 2.1: (a) an analog representation and (b) a symbolic (and propositional) representation of the relative location of two buildings.

When considering representations of space, it is important to consider their properties. Obviously, a verbal description of an environment has different properties than a map. What makes them different? Figure 2.1 shows two examples for representations. The relevant properties of representation (a) are similar to that of map. Likewise, the relevant properties of representation (b) are similar to that of a verbal description. Both representations, (a) and (b), stand for the relative location of the town hall and the church.[4] The relative location is represented by virtue of certain relationships. In order to understand the representational configuration, it is crucial to understand the relationships between the representation and the represented world.

Representation (a) stands for the relation in a pictorial, map-like way (ignoring—for now—that the names of the buildings are represented in textual form; I come back to this point in Section 2.2.2). This type of representation is called 'analog representation' (e.g., Palmer, 1978). Such representations are often described as picture-like and are also

[4]As discussed in Section 2.1, representations stand for one or more aspects of the representing world. As most analog representations, representation (a) potentially stands for multiple aspects of the represented world. For example, it can represent the relative location of the buildings and the sizes of the outlines of the buildings. In the discussion, I concentrate on the representation of relative location.

called ‘analogue’, ‘diagrammatic’, ‘isomorphic’, ‘iconic’, or ‘depictive’ representations. The relationship between representation (a) and the represented world is as follows: The relative location of the two buildings is represented by the relative location as depicted in the representation. The representing world shares the property to be spatial with the represented world. It shares the constraints on relative locations between buildings that are inherent in the represented world. For example, the town hall cannot be to the right of the church and to the left of the church at the same time (given there is only one town hall and one church).

Analog representations of space can be defined in a sense called ‘strict’ when the definition includes that the representation itself is spatial and information is preserved in a way that it resembles that which it stands for (Palmer, 1978; Quinlan & Dyson, 2008). According to the less strict view, an analog representation does not need to be image-like and does not need to resemble the represented world (e.g., Palmer, 1978; McNamara, 1999). However, as discussed by Palmer (1978), an analog representation has to preserve some aspect of the structure of the represented world (otherwise, the two classes of representations are not oppositions anymore). I follow Palmer’s definition of analog representations.

Representation (b) stands for the relative location of the buildings in a different way than representation (a) does. The spatial relation between two the buildings is represented by the following combination of symbols: is_left_of(town hall, church). Such a representation is called a ‘symbolic representation’. Symbolic representations are called ‘non-analogue’ or ‘sentential’ representations, as well (Bernsen, 1994; Acartürk, 2009). Symbolic representations are representations in which objects are represented by arbitrary symbols (i.e., the relation to the represented world is not based on structural correspondence). In representation (b), the two buildings are represented by the object symbols town hall and church. Relations between the objects of the represented world are represented by additional relational symbols that stand for the relations between the objects (Palmer, 1978). The predicate is_left_of fulfills the task to represent the relative location between the buildings in representation (b). The representation allows for the construction of propositions; therefore, representation (b) is a propositional symbolic representation. Propositional representations are a special class of symbolic representations: They allow for the construction of statements that have a truth value, such as is_left_of(town hall, church) in representation (b). Other examples for propositional (and, consequently, symbolic) representations are natural language and formal logic systems such as the predicate calculus (Palmer, 1978).

2.2.2 Hybrid Representations

Most representations are more complex than the discussion in the last section suggests. Representations refer to aspects of the represented world, and can refer to more than one aspect or to one aspect in multiple ways. Representations can be combinations of analog and symbolic representations. This becomes apparent when representation (a) (Figure 2.1) is reconsidered. The representation of relative location, in line with the discussion above, is an analog representation. Yet, the buildings are identified with written labels stating their names. Obviously, this naming of the buildings is based on a symbolic representation: The names of the buildings are represented in a textual form. In this respect, representation (a) is a symbolic representation (it is not a propositional representation because it does not allow to construct propositions). Still, the relative spatial location is represented in an analog way. The representation combines two classes of representations: an analog and a symbolic one. In the following, this class of combined representation is called 'hybrid representation' (the term 'hybrid representation' is used in a similar meaning in Habel, Pribbenow, & Simmons, 1995; Habel, 1996, 2003b; Lohmann et al., 2010). Usually, maps are hybrid representations that make—in the case of visual maps—use of written labels to establish correspondence to real-world objects based on propositional representations.

I exemplified hybrid representations (i.e., combinations of analog and symbolic representations) by discussing maps. Of course, maps are not the only case of hybrid representations. For instance, movies that combine language and pictorial means, graphs with written labels, and paintings and photographs with titles are hybrid representations. Hybrid representation is a standard case for external representations rather than an exceptional case.

2.2.3 Sensory and Representational Multi-Modality: A Definition

External representations can be perceived via different sensory channels such as audition and vision, which is one type of multi-modality called 'sensory multi-modality'. In addition, they can be combinations of specific representational systems, such as German or English natural language and (visual) maps. The latter type of multi-modality is called 'representational multi-modality' (Bernsen, 1994). How this form of multi-modality should be defined is controversially discussed in a number of disciplines. The most systematical analyses were made in linguistics and media philosophy. A common meaning currently has not been established (Fricke, 2008; Schneider & Stöckl, 2011). In the following discussion of multi-modality, I borrow concepts from linguistics and media philosophy. I do not adapt to the terminology used in these disciplines. Rather, I use a terminology that is consistent with that used in the last sections (in which the terminology was mainly based on that used in cognitive science).

In order to discuss multi-modality and multi-modal representations, it is important to first consider what a modality is. As mentioned earlier, in one sense, the term 'modality' refers to a sensory channel such as audition or vision. For example, imagine a speaker A giving directions to a recipient B having a road map at hand: 'Here, go to the left.' While A produces the utterance, he or she points with a finger at the intersection (on the road map) where B has to turn. The representation is 'sensory multi-modal': The directions are perceived via a combination of different sensory channels, namely audition and vision (by observing the pointing gesture and listening to the utterances).

However, the directions in the example do not only have the property to be perceived via different sensory channels, in addition, different *representational systems* are

Table 2.1: The layers of a representation.

Layer	Distinguishing Feature	Examples
3	Representational system(s)	Written English language, road map, ...
2	Class of representation	Analog, propositional, or hybrid
1	Material substrate	Print technology, sound waves in the form of speech, ...

at work.[5,6] A representational system can be circumscribed as 'the code at work in' the representational configuration. For example, in representation (b) discussed in Section 2.2.1, the representational system consists of the object symbols town hall and church and the relation symbol is_left_of and knowledge about the interpretation and combination possibilities of the symbols. I use a very broad notion of representational systems[7]: the stand-for relation of all types of representations is based on representational systems of one kind or another. For example, like language, pictorial representations (such as representation (a) discussed in Section 2.2.1) mean something by virtue of representational systems.

Representations can combine different representational systems, as the example of the directions with a map discussed earlier shows. Consequently, this example is not only an example of sensory multi-modality, but also of representational multi-modality. The different forms of multi-modality can be further explicated: They refer to different layers of a representation. A representation can be regarded as consisting of three levels, which are shown in Table 2.1. The first layer, called 'material substrate' (I adopted this term from Stetter (2005), who uses the German term 'materielles Substrat') is the physical framework in which the representation is realized. For example, the most important material substrate of this thesis is the print (the paper, the ink, and the corresponding technology) or, in case its read on a computer screen, it is the electronics, which allow you to see my writing on the screen. If humans speak, the material substrate is the air surrounding them, which transports the sound waves, and also the vocal chords and tongue and the perceptive functions allowing the addressee to hear the words. A material substrate is the necessary condition of each external representation: Air and vocal chords are needed to speak, without them, it is not possible to speak.

[5]I use the term 'representational system' with the same meaning like the term 'symbol systems' as discussed by Goodman (1976). To avoid confusion with the term 'symbolic representations', I use the term 'representational system' in this work. I do not provide a detailed definition of representational systems here, but refer the interested reader to Goodman's definition of symbol systems (see also Schneider & Stöckl, 2011).

[6]The term 'representational system' has the same meaning as the term 'medium' in what is called a 'code-oriented' sense. In a code-oriented sense, media are are distinguished by the rules with which a user of a medium associates meaning to the external representation (Posner, 1985; Fricke, 2008).

[7]This broad notion of the term is consistent with Goodman's (1976) discussion of symbol systems. A discussion can also be found in Lohmann (2009).

The second layer is the layer of the class of the representation; that is, how and on which representational basis the representation is constructed in general. The classes of representations that were discussed in Sections 2.2.1 and 2.2.2—analog, propositional, and hybrid—describe representations on this layer of analysis.

The third layer is more specific than the second layer; it is the specific representational system of an external representation. The third layer is characterized by the representational system or systems of the representation. For example, the directions discussed above are combinations of spoken (English) natural language and a road map.

Now, the terms that are needed for a proper discussion of the different types of multi-modality are at hand: Sensory multi-modality refers to representations combining material substrates that are perceived via different sensory channels. For example, the directions during which the speaker points with on a location on a road map with his finger are sensory multi-modal.

In addition, the directions incorporate two different representational systems, namely, the road map and a specific natural language (say, English). Therefore, the directions are an example of representational multi-modality, as well. Whether or not a representation can be characterized as representationally multi-modal depends on whether different representational systems are at work. Representational multi-modality is established on the third layer of a representation (but cf. Acartürk, 2009).

Especially for accessible technology, representations in which the use of one sensory modality is limited have to be developed (Loomis et al., 2012). In some cases, the change of the sensory modality (by changing the material substrate) is applicable. For example, speech synthesis is used to make text accessible, without important loss of information.[8] However, this substitution is not possible for all external representations: Graphics cannot be transformed to speech without losing potentially important information. The same problem exists in the case of maps. To increase accessibility of maps, multi-modal representations were developed in which tactile information is enriched by audio information (see Section 2.3.3). While these representations are sensory multi-modal, they are only in a weak form representationally multi-modal because the use of natural language is limited. Mostly, it is restricted to uttering the names of the objects on the map. Due to the need for an integration of sequentially perceived knowledge on tactile maps (see Section 1), this only reduces a part of the problem.

VAVETaMs extend the representational multi-modality of audio-tactile maps by utilizing assisting utterances that, for example, include statements about the relative location of map objects. By this, VAVETaMs benefit from representational multi-modality. Representational multi-modality is an option to overcome problems of simple sensory substitution (i.e., providing the same information via another sensory channel) such as limited capacity of the substituting sensory channel.

[8]This is not to say there is no loss of information when text is synthesized. There is loss of information because the characteristics of two dimensional text—such as typesetting—are lost.

2.3 Maps and Language

2.3.1 (Virtual) Tactile Maps

Humans use maps in a multitude of scenarios. Whereas knowledge of visual maps is acquired by using the visual sense, tactile maps and virtual tactile maps are explored by touch. This sense consists of two subsystems, cutaneous and kinesthetic perception (Lederman & Klatzky, 2009). Cutaneous perception is enabled by mechanoreceptors and thermoreceptors in the human skin. It allows to perceive surface information such as the texture of an object. Information for kinesthetic perception comes from mechanoreceptors that are embedded in muscles, tendons, and joints. The kinesthetic perceptual system is also called the 'proprioceptive system'. Usually, both inputs are combined in haptic perception.

Tactile maps are used as substitutes for visual maps by visually impaired people. As spatial representations, they can provide spatial information like that represented in visual maps. Compared to visual maps, tactile maps have drawbacks in terms of speed and accuracy of the knowledge acquisition process. This results from the sequential perception of the haptic sense: Tactile maps have to be explored part by part and the knowledge acquired has to be integrated to form consistent spatial knowledge (Loomis et al., 1991; Jacobson, 1998b). Furthermore, using text for names of objects is problematic in representations for visually impaired users. Braille labels are too inflexible and only understood by 15–25 percent of blind people (Lötzsch, 1994; Jacobson, 1998b).

Current devices for human-computer interaction allow for haptic interaction using a pen-like handle. With these devices, virtual haptic objects can be perceived. Virtual haptic objects are three-dimensional objects that can be explored by the device. They are called 'virtual' because their presentation with the device is based on data and this is the only form in which the objects exist (like objects in visual virtual environments). With respect to their modeling, these objects are similar to the objects that are visually rendered in visual virtual environments (e.g., 3D computer games). Because the virtual objects are explored with a handle, the cutaneous information that can be used to gather information about the objects to be explored is relatively sparse. Often, the direct exploration of surfaces providing cutaneous feedback is called tactile perception and the virtual variant that mainly relies on kinesthetic information is called haptic perception. However, this distinction is not straightforward, because, as discussed, the haptic perceptual system includes both, cutaneous and kinesthetic perception. The perception of the virtually realized variants of tactile maps discussed in this thesis is mainly (though not only) based on kinesthetic perception.

Before the advance of computer technology and devices for haptic human computer interaction, tactile maps had to be handcrafted (Tatham, 2003). The manual production of tactile maps is costly and even minor edits—for example, due to changing environments—result in major efforts. With the advance of computer technology, techniques that enable tactile printouts have been developed. For example, special papers (called 'swell papers' or 'microcapsule papers') that can be printed with a common ink-jet printer were developed. Furthermore, tactile printers were developed that mechanically emboss card-board sheets. Printout techniques are often used in combination with touch pads that provide information which part the user explores to create audio-tactile maps (see Section 2.3.3). In comparison to virtual tactile maps, systems based on printout techniques have the drawback that printouts have to be produced, which makes them less flexible than systems that are based on a device such as the VAVETaM system.

The Phantom device, which we use to present virtual tactile maps, was initially based on a thimble-like interface design (Massie, 1993, see also Massie & Salisbury, 1994). The original design of the device allowed for three degrees of freedom. Its abilities to ease human computer interaction for visually impaired people were investigated by Sjöström (2001a, 2001b), resulting in a couple of general design guidelines, such as to include well defined and easy to find reference points. Later, the Phantom devices were changed to systems with pen-like handles and six degrees of freedom. Figure 2.2 shows a current, commercially available product, the Sensable Phantom Omni. This device is used in the research described in this thesis. It has six degrees of freedom, of which three are with force feedback. The spatial resolution is appropriate for the presentation of *virtual tactile maps*, which are virtual-environment counterparts to traditional physical tactile maps.

Figure 2.2: The Sensable Phantom Omni interface.

The pen-like handle of the Phantom Omni device is attached to a moveable arm. This arm can be thought of as a reverse robotic arm. Sensors register the position of the point where the axis of the handle and the arm meet. This point is called the ‘interface point’ (Salisbury, Brock, Massie, Swarup, & Zilles, 1995). Servomotors in the device enable the perception of virtual objects by generating appropriate forces. A 3D model of the virtual object that is stored on a connected computer is used to calculate appropriate force depending on the current position of the interface point. When the user moves the handle towards an area that is modeled as solid-object, a force is generated. As the device generates force depending on the position of the interface point, a virtual haptic perception is enabled.

The user can explore and feel virtual objects with impenetrable surfaces using the device. To explore these virtual objects, the user moves the handle of the device. It allows movements in all three spatial directions. Compared to real haptic perception, this form of interaction has the drawback of limiting interaction to the exploration of three-dimensional models with the pen-like handle instead of one or two hands. The haptic interaction with virtual objects can be described as one-point interaction. Therefore, cutaneous perception is comparably sparse. Furthermore, it is not based on direct contact to the explored object and, hence, some information that is available in real-world haptic perception, such as information about temperature of the object, cannot be perceived.

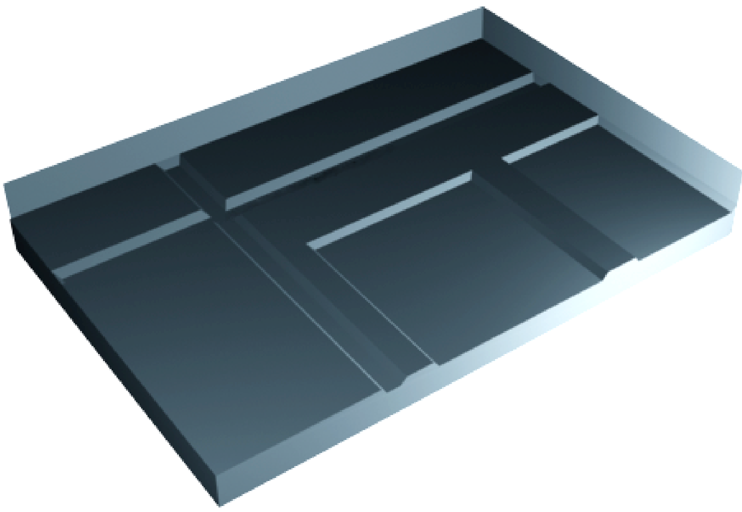


Figure 2.3: A visualization of a cross section through a virtual tactile map.

The virtual objects modeled can be tactile maps in which streets and potential landmarks, such as buildings, are marked as indentations. A cross section through such a virtual tactile map is shown in Figure 2.3. The walls shown in the back of the figure are part of a virtual box surrounding the map. The walls of the surrounding box are realized such that when the model is explored with the device, they are only solid when approached from the inner space of the box. Therefore, the surrounding box can be entered and the map located in it can be explored but unintentionally leaving the map is prevented. Users explore the virtual tactile maps by performing movements on the surface of the map and especially by following grooves representing streets and exploring the other map objects represented by two-dimensional indentations (i.e., regions lower than the base plane) with movements that enable them to detect the shape of the objects.

In their sequential exploration, they focus haptically on objects or parts of objects about which they wish to acquire knowledge; for the respective dynamically determined region I use the term *'haptic focus'* in the following (Lohmann et al., 2011). Mostly, an object of a map, such as a street is in the haptic focus.

As discussed earlier, the Phantom allows for one-point interaction. Possibly, performance could be increased by large-scale raised-dot tactile displays that allow for haptic exploration with two hands. Such devices are used in research projects (e.g., Zeng & Weber, 2010), but they are expensive.[9] In case such devices get available in lower price ranges, the technical solution described in this thesis could likely be adapted because it relies on exploration categories rather then on detecting when objects are touched. Therefore, it would be possible to identify and react to typical exploration movements in such displays when suitable categories of exploration movements (such as following to parallel streets with two fingers or using one hand as reference to estimated distances between objects) and corresponding helpful verbal information would be developed. In contrast to large-scale tactile displays, the Phantom is commercially available and comparably low priced.

2.3.2 Reference Systems in Spatial Language and Maps

Humans can ask other humans for route directions. If they are successful, they receive verbal descriptions of routes that lead them to their goal. Verbal directions can be effective means that ease or enable navigation. Their structure and constraints on their effectiveness have been investigated (e.g., Daniel, Ariane, Manghi, & Denis, 2003; Meilinger, 2005; Hund, Haney, & Seanor, 2008; Tenbrink & Winter, 2009). Language can, of course, also be used to describe other spatial scenarios. For example, schematic maps of indoor spaces and road maps can be described using natural language (Taylor & Tversky, 1996).

Basically, natural language enables communication of spatial information in two ways: identification of objects and communicating (relative) positions of the objects identified (Landau & Jackendoff, 1993). To solve the first task, human language offers a set of nouns and names. With respect to spatial properties, for instance, the shape of an object can be used to identify it by making use of shape expressions such as 'square', 'circle', or 'cube'. To solve the second task, spatial prepositions such as 'above', 'below', 'on', or 'right of' can be used. For example, humans can say 'The laptop is on my desk' and use the preposition 'on' to refer to the relative location of laptop and desk.

A spatial preposition relates one object to another. It relates the object to be located to a reference object. The object to be located is called 'figure'. The object that serves as reference is called 'reference object'. Spatial prepositions define a region in which the figure is located—for example, the space 'on my desk' (Landau & Jackendoff, 1993; Talmy, 2000).

[9] Recently, the devices were made available by Metec (http://www.metec-ag.de). Currently, they are too expensive for individual use (about 50 000 Euro including software, information from Metec).

In order to understand utterances that make use of spatial prepositions, the addressees of the utterance must understand the producer's use of reference systems.[10] Reference systems are part of the 'code at work' in representations, as such, they are part of the representational system.

As discussed by Eschenbach (1999), reference systems are "structures that support spatial cognition" (p. 329), which are essential for the interpretation of spatial prepositions. A reference system is a relational system that is based on the related objects, the reference object(s), and the spatial relations that can hold between them (Shelton & McNamara, 2001).[11] Some reference systems additionally include a viewer (usually, the addressee of an utterance). Utterance (1) is an example for an utterance, which is ambiguous with respect to the reference frame adopted.

(1) The boy is in front of the house.

Depending on the context, this utterance can have several meanings: The boy can be in front of the house (i.e., where the door is), or the boy can be between the house and the speaker and the addressee (see Figure 2.4, for a visualization).

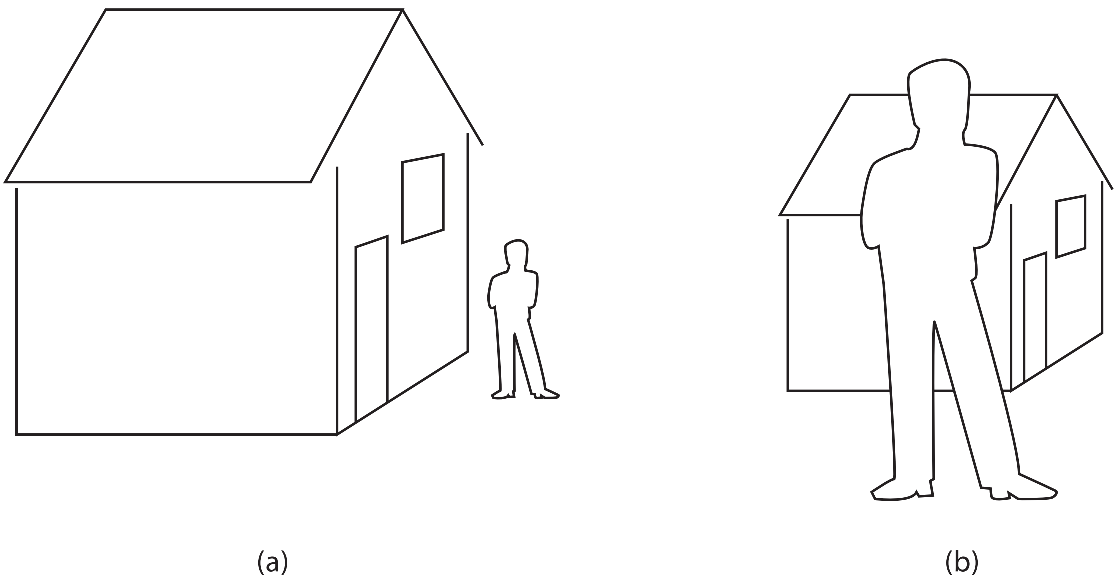


Figure 2.4: Visualization of the ambiguous meaning of the preposition 'in front of'.

[10] Eschenbach (1999) suggests that the term 'reference system' is used for a concrete frame of reference established by a collection of objects. According to Eschenbach's terminology, which is adopted here, the term 'frame of reference' refers a more abstract level of geometric entities that model the geometric structure of reference systems.

[11] As Eschenbach (1999) discusses, reference systems are not coordinate systems or based on coordinate systems, as stated in or implied by the description by other authors (e.g. Levelt, 1989; Carlson-Radvansky & Irwin, 1993; Levinson, 1996b, 1996a).

As the example shows, humans can make use of different reference systems when they talk about space. Different classifications of reference systems exist (Newcombe & Huttenlocher, 2003; examples for systems are discussed in Carlson-Radvansky & Irwin, 1993; Levelt, 1989; Taylor & Tversky, 1996; Shelton & McNamara, 2001). Levinson (1996a) describes three kinds of reference systems: intrinsic, relative, and absolute (see Eschenbach, 1999, for a formalization of these reference systems that does not rely on coordinate systems). When an intrinsic reference system is used in communication, the figure object is located with respect to inherent features or properties of the reference object. For example, utterance (1) can be understood in terms of an intrinsic reference system. The house has a front, which is an inherent feature of it. Figure 2.4 (a) shows a constellation that corresponds to the described interpretation of utterance (1).

In verbal directions, a frequently used reference system is intrinsic reference based on the movement direction (or body orientation, usually, these reference systems are the same) of the addressee. For example, a direction can be ‘Follow the road till you reach a T-intersection. Turn to the left and follow the street until you see a church. Behind this church, turn left again.’ The two turning instructions are formulated with respect to the anticipated movement direction of the traveller. Therefore, this frame of reference is also called ‘egocentric’ frame of reference (Taylor & Tversky, 1992b, 1996).

The interpretation of utterance (1) with respect to a relative reference system corresponds to Figure 2.4 (b). In this interpretation, the utterance expresses a ternary relation between a perceiver’s location, the boy, and the house. Usually, the addressee’s location is the same like the speaker’s location. Therefore, the boy is between the house and the location of the addressee and the speaker.

Absolute reference systems do not depend on an addressee’s location or a feature of an object. Rather, they are established by a feature of the environment or they are of conventional nature. For instance, gravity establishes such an environmental absolute system (Levinson, 1996a). Another paradigm example for an absolute reference system are the cardinal directions ‘north’, ‘east’, ‘south’, and ‘west’, which are of conventional nature. Absolute reference systems remain invariant to movements of the speaker or the addressee. For example, the sentence ‘Hamburg is north of Vienna’ remains true independent of the location of the speaker and the addressee.

As structures that support spatial cognition, reference systems play a role in maps as well. For example, many maps adopt the cardinal direction system, as indicated by the compass card on the map. In addition, a reference system is introduced by the two-dimensional representation on paper, which has a fixed orientation. This orientation is indicated by the text direction of the written names of objects on the map, the legend, and so on. Objects on maps stand in spatial relationships to each other. Therefore, it is possible to say that an object on a map is, for instance, left of another object. It is also possible to state that an object is above another entity, even though maps are often read lying flat on a table. This form of reference is similar to that used to refer forth and back in written texts. It is used in statements such as ‘see above for a definition’. If a map is oriented such that north is upwards, obviously, up and north, east and right, and so on, fall together. However, not all maps are oriented in this manner. When relative spatial locations on a map are described while the map is perceptually accessible (such as giving descriptions with a map pointing at locations on it during description, to pick up the example discussed in Section 2.2.3) the reference systems have to be coordinated. With respect to the relative location of fixed objects, a speaker can choose among the two reference systems relevant for maps: it is possible to say that a building is left of another building and it is possible to say that it is west of another building (though these two statements only correspond in meaning in maps where the north direction falls together with the upwards direction).

Figure 2.5: A situation in which reference systems conflict.

As the example of the description having a map at hand shows, in addition to the reference systems that are introduced by the map itself, a potential movement of the addressee creates a reference system intrinsic to the movement direction (and anticipated body orientation of the addressee). Figure 2.5 shows such a situation. The figure shows a part of a street network depicted as in a map. If directions are given that involve the movement from the vertically shown street towards the T-intersection ('down' in the figure), the two absolute reference system based on map orientation and the intrinsic systems based on the movement direction conflict. 'Left' with respect to the movement direction is 'right' with respect to the map orientation.

Because the haptic exploration of virtual tactile maps involves movements of the hand, it could also be the case that intrinsic systems based on movement directions should be used in assisting verbalizations. However, most map users do not explore the virtual variants of tactile maps with steady hand movements, but rather move back and forth when they focus on a map object. In a corpus study we conducted (see Section 3.2), absolute reference frames were used by participants who were asked to verbally assist an explorer of a virtual tactile map. Therefore, the assisting utterances that were positively evaluated in the experiments reported in Chapter 3 make use of an absolute reference system.

Additionally to the capacity of spatial language discussed, humans can use language to express information about geometric relations that objects have. For example, humans can produce utterances such as '42$^{\text{nd}}$ Avenue is parallel to 36$^{\text{th}}$ Avenue'. In a corpus study, we found that humans who are given the task to verbally assist an explorer of a tactile map make use of this capability of natural language (see Section 3.2). As the perception of parallelism is problematic with the one-point interface, providing such information via assisting utterances is potentially useful. There exist more geometrical terms to describe spatial settings. For example, it is possible to describe the angle in which two streets meet as ‘approximately right angle’ or to describe it as ‘41 degrees angle’. Yet, humans who verbally assist tactile-map explorers do only very rarely (1 utterance in 1011 collected utterances in the corpus study reported in Section 3.2) make use of this capacity of natural language (probably, because the verbal assisting utterances are time critical and utterances of this complexity are difficult to produce and to understand).

2.3.3 Audio-Tactile Maps

> “A tactile map is just a jumble of patterns and lines, now it talks to me, I can explore at my own pace, finding my own way.” (a blind user of an audio-tactile map as quoted in Jacobson & Kitchin, 1997, p. 326)

Tactile maps and virtual tactile maps have to be explored part by part. The knowledge acquired has to be integrated to form consistent spatial knowledge (Loomis et al., 1991; Jacobson, 1998b). Furthermore, Braille text is too inflexible to use it to present names of streets and other objects and only understood by 15–25 percent of blind people (Lötzsch, 1994; Jacobson, 1998b). To combine tactile maps with natural language is a promising way to facilitate the knowledge acquisition from tactile maps (de Almeida (Vasconcellos) & Tsuji, 2005; Rice, Jacobson, Golledge, & Jones, 2005; Golledge, Rice, & Jacobson, 2005; Paneels & Roberts, 2010; Buzzi, Buzzi, Leporini, & Martusciello, 2011). In this section, I review approaches towards audio-tactile maps. I center the review on systems using the Phantom device in combination with sound or natural language output. Though, I review some approaches with other forms of haptic map presentation when they are relevant.

The existing systems use natural language to present the names of objects, but in contrast to VAVETaMs, natural language is not (or very sparsely) used to communicate spatial information. The first audio-tactile map system was developed by Parkes (Parkes, 1988; de Almeida (Vasconcellos) & Tsuji, 2005). In this system, a physical tactile map is placed on a special touchpad called ‘NOMAD’. This touchpad enables the association of sound or speech to objects. Several researchers further developed this idea—see Table 2.3.3, for an overview, and Wang, Li, Hedgpeth, and Haven (2009), for a recent system.

Different approaches were made in the direction of more flexible systems that use computer devices to present tactile maps and that do not need physical tactile-map overlays. For example, systems were developed that test the substitution of haptic feedback by audio-only feedback (e.g., Krueger & Gilden, 1997). Furthermore, systems that use conventional gaming hardware were developed (e.g., Lahav & Mioduser, 2000).

Research has shown that the Phantom device can successfully be used to present virtual tactile maps. Moustakas, Nikolakis, Kostopoulos, Tzovaras, and Strintzis (2007) describe the use of a Phantom device to create virtual tactile maps combined with audio output. The system was successfully evaluated in a user study, as well. A similar system was developed by De Felice et al. (2007). The system was successfully tested in a user study. Recently, Kaklanis et al. (2011) used the Phantom to present virtual tactile maps that are automatically generated from geographic information available online. In conclusion, the literature shows shows that Phantom devices can be used to present virtual tactile maps.

Additional evidence for the usability of the Phantom comes from applications for visually impaired people that do not use map-like representations, but use the Phantom such that the interaction is comparable to exploring environments with a cane. Magnusson and Rassmus-Gröhn (2004) tested such an environment. The environment models a traffic scenario. It includes moving objects such as cars and is based on a combination of sound and haptics. Lahav, Schloerb, Kumar, and Srinivasan (2012) successfully used the Phantom device to present an indoor environment, also using a cane-like interaction.

With respect to the evaluation of audio-tactile systems in general, mostly, informal studies and user studies have been conducted (see Table 2.3.3). In conclusion, empirical research shows that audio-tactile maps are usable and understandable means for the acquisition of spatial knowledge for visually impaired people. A few controlled experiments evaluating the effectiveness of the knowledge acquisition process of audio-tactile maps—both, based on virtual environments and on physical tactile maps—have been conducted.

In the two experiments discussed in the following, it was compared whether knowledge acquisition using an audio-tactile map is more effective than direct experience. The results indicate that audio-tactile maps are more effective than direct experience, providing support for the effectiveness of VAVETaMs.

In an experiment reported by Jacobson (1996), environmental knowledge acquisition with an audio-tactile map based on NOMAD was compared to direct experience. In the map-learning condition, the environment was enhanced with special artificial audio emitters ('audio beacons'). Participants were able to produce better route sketches when they used the audio-tactile map then when they learned the route by direct experience.

In an experiment reported by Lahav and Mioduser (2008a, 2008b), it was compared whether spatial-knowledge acquisition of an indoor environment by an audio-tactile representation is more effective than knowledge acquisition by direct experience. The virtual audio-tactile environment combined audio-output with haptic feedback from a force feedback joystick (Lahav & Mioduser, 2000). In the experiment, the system was used to present a room with a relatively simple setup of objects contained.

With respect to accuracy of spatial knowledge, the experimental group, which entered the room for the first time, outperformed the control group, which had learned the layout of the room by direct experience. The experiment shows that virtual, audio-tactile environments can be successfully used as means to acquire spatial knowledge of a real-world space. Spatial knowledge acquired in this way eases actual navigational tasks in the real-world environments, which supports the idea of VAVETaMs.

Other researchers have evaluated audio-tactile approaches in terms of usability and user acceptance. Interfaces that try to substitute haptic feedback with sound were positively evaluated (Krueger & Gilden, 1997; Jacobson, 1998b). In these interfaces, users explore a common touch pad with one finger and objects on the maps are indicated by a sound that is emitted when the respective area on the map is touched. These interfaces share a perceptual constraint with VAVETaMs: Haptic interaction occurs only at one point and the cutaneous feedback is sparse.

The Hyperbraille research project uses a large-scale Braille display to present audio-tactile maps (Zeng & Weber, 2010, 2012; Schmitz & Ertl, 2012). The Braille display was developed in a research project and is too expensive for individual use (as discussed in Section 2.3.1).

The question whether and how assisting utterances can be used to further increase the effectiveness of audio-tactile maps was not yet addressed in controlled experiments. The systems that are described all associate verbal or audio output to areas or objects on the (virtual) tactile map. If the users either touch or click on (e.g., Miele, Landau, & Gilden, 2006) an object to which verbal our audio information is associated, the corresponding verbal or audio output is started. The output is started regardless of whether information has been given before or *how* the user explores the object (e.g., how long the object is explored). The systems discussed work on the basis of what can be described as 'static trigger regions'. Fixed verbal or audio output is associated to an area or object on the map. Natural language is only used to state the names of the objects to enable their identification. There are few examples for systems that include verbal information about the relative spatial location of objects: Jacobson (1998a) discusses a system that provides a short description of the map presented when a button is clicked. Petrie et al. (1996) also briefly mention that verbal descriptions could be given in an audio-tactile system, but do not discuss the topic in detail or provide an evaluation of using such descriptions.

As a corpus study discussed in Section 3.2 showed, if humans are asked to assist a user of a virtual tactile map, they make use of spatial language to a much greater extend than informing the map user about the names of objects. The goal of this thesis is to investigate the prospects of automatically generating assisting utterances similar to that which a human assistant gives and thus, to extend the successful approaches towards audio-tactile maps that were discussed.

Table 2.2: Evaluation of audio-tactile map systems for visually impaired people (ordered by publication date).

Publication, System Name	**Haptic or Tactile Presentation**	**Audio or Verbal Output**	**Evaluation**
Parkes (1988, 1994), NOMAD	Printout map placed on touchpad	Sound for objects and verbalization of distance between two points	
...	...	...	...

Publication, System Name	**Haptic or Tactile Presentation**	**Audio or Verbal Output**	**Evaluation**
...	...	...	...
Lötzsch (1994), AUDIO-TOUCH	Printout map on touch pad	Speech synthesis	
D. G. Evans and Blenkhorn (1995), Talking Tactile Map (TTM)	Physical tactile map on touch pad	Speech synthesizer	
Petrie et al. (1996), MOBIC Pre-Journey System (MOPS)	Flexible, e.g., physical map on touch pad	Speech synthesis (describing user's position)	
Jacobson (1996)	Based on NOMAD	Auditory icons (non-verbal sounds)	Evaluated with eight visually impaired participants asked to learn a route through an environment presented with the audio-tactile map.
Jacobson and Kitchin (1997), Hypermap	Touchpad without tactile feedback	Speech synthesis	
Krueger and Gilden (1997), KnowWhere	No tactile feedback	Names or abbreviations and various sounds	Informally evaluated with five blind participants.
Hudson (1998)	Based on NOMAD	Names of objects via speech synthesis	
Jacobson (1998a, 1998b)	Touch pad without tactile feedback		The system was evaluated for usability with ten visually impaired participants.
...	...	...	...

Publication, System Name	**Haptic or Tactile Presentation**	**Audio or Verbal Output**	**Evaluation**
...	...	...	...
Lahav and Mioduser (2000, 2003)	Force-feedback joystick	Auditory feedback (no detailed description included)	Evaluated by two visually impaired participants. See Lahav and Mioduser (2004, 2008a, 2008b) (following in this table) for an evaluation with more participants.
Parente and Bishop (2003), BATS	Force-feedback joystick (and other hardware)	Speech synthesis and sounds	
Siekierska and Müller (2003); Siekierska, McCurdy, and Peterson (2008)	Physical tactile map on a touchpad	Prerecorded verbal descriptions or environmental sounds	
Magnusson and Rassmus-Gröhn (2004, 2005)	Sensable Phantom	3D sound	Eight of ten participants were able to handle the system.
Lahav and Mioduser (2004, 2008a, 2008b)	Force-feedback joystick	Virtual footstep sounds, names of objects	Empirical evaluation with 31 visually impaired participants.
Miele et al. (2006), Talking TMAP	Physical tactile map on touch pad	Speech synthesis	
De Felice et al. (2007), Omero	Sensable Phantom Desktop	Sounds or speech synthesis	Evaluation for usability with eight visually impaired participants.
Kostopoulos et al. (2007) and Moustakas et al. (2007)	Sensable Phantom Desktop	Speech synthesis of street names	Evaluation for usability with 19 blind participants (Moustakas et al., 2007).
Lawrence, Martinelli, and Nehmer (2009)	Haptic mouse	Sound	Informal user testing with visually impaired participants.
...	...	...	...

Publication, System Name	**Haptic or Tactile Presentation**	**Audio or Verbal Output**	**Evaluation**
...	...	...	...
Wang et al. (2009), Paladugu, Wang, and Li (2010)	Printout map on touch pad	Names of objects	Usability evaluation with six blind participants.
Zeng and Weber (2010, 2012) and Schmitz and Ertl (2012)	Large-scale braille display	Names of objects	User study with four blind users.
Kaklanis et al. (2011), HapticRiaMaps	Phantom	Sonification	Informal evaluation with ten visually impaired participants.
Lahav et al. (2012), BlindAid	Phantom	Spatial Audio	The system was evaluated with four blind participants.

2.4 Humans' Spatial Knowledge

2.4.1 Spatial Mental Models

The goal of external representations of space is to enable acquisition of knowledge about the represented spatial environment. In order to discuss how spatial knowledge is represented by humans, it is necessary to discuss which kind of space is in the focus of interest. The term 'space' refers to very differently sized and structured environments. Montello (1993) suggests that spaces can be classified by their size into four categories (see also Kitchin & Blades, 2002, for an overview of classifications of space). The categories suggested are: (1) figural space, which is smaller than the body and can be directly perceived; (2) vista space, which is larger than figural space and can be perceived without (a lot) of locomotion; (3) environmental space, whose perception needs locomotion; and (4) geographical space, which is larger than environmental space and whose understanding is only possible by using representational means such as maps. The kind of space that corresponds to the focus of my research is environmental space, the space in which buildings, neighborhoods, and cities are organized (Montello, 1993).

Humans' internal spatial representations of environmental space are often called 'cognitive maps'. The term has been introduced by Tolman (1948) (to describe the internal representation of space of rats). Since its introduction, the term has been the subject of discussion. It was often challenged for its implication to be image-like. Yet, it is continuously used as a metaphor in spatial-cognition research (e.g., Downs & Stea, 1977; G. W. Evans & Pezdek, 1980; Kitchin & Blades, 2002). Other terms are used to describe humans' internal representations of space, most of them with similar implications. Among them are 'environmental image' (Lynch, 1960) and 'mental map' (Beck & Wood, 1976; see Kitchin & Blades, 2002, for another overview of terms used). Some researchers avoided terms that imply an image-like representation: Siegel and White (1975) use the term 'spatial representation' and Appleyard (1969) uses the term 'urban knowledge'. A major argument against the term 'cognitive map' is that its implication to be image-like suggests a purely analog representation in the strict sense (supposing an isomorphic representation with a 1-to-1 mapping of spatial information and representation; see Section 2.2 and, e.g., Kuipers, 1982). Therefore, this term can easily lead to misunderstandings. As suggested by Tversky (1993) I use the term 'spatial mental model' to refer to humans' internal representation of survey knowledge.

According to Tversky, a major argument against the term 'cognitive map' is that humans acquire separate pieces of knowledge when they interact with an environment, which are not necessarily integrated in a coherent representation. Such pieces of knowledge are—for example, recollections of journeys, memories of maps, and recalls of verbal directions. The different pieces of knowledge can lead to various forms of spatial knowledge, which do not need to be map-like in general. Tversky suggests to call spatial knowledge that is not organized in a coherent whole 'cognitive collage'. Yet, simple or well-learned environments can be represented in coherent internal spatial representations. In this case, spatial mental models are built up. These representations allow for perspective-taking, reorientations, and spatial inferences and they represent coarse spatial relations appropriately, which distinguishes them from cognitive collages (Tversky, 1993).

However, unlike true maps, spatial mental models do not preserve metric information. Rather, various spatial aspects are systematically distorted (but cf. Newcombe & Huttenlocher, 2003, who discuss that internal representations can be metric and still show systematic distortions). For example, a systematic distortion was disclosed in a study in which students from San Diego were asked to sketch the direction between San Diego in California (at the west coast of the USA) and Reno in Nevada (inland) (Stevens & Coupe, 1978). The students' sketches showed Reno west of San Diego, which is wrong. Stevens and Coupe (1978) ascribe that finding to an hierarchical representation of space. Instead of a purely analog, metric representation, humans remember the states the cities are part of, and the relative locations of states. The locations of the cities are inferred from the relative locations of states, which leads to the distortion described. There is a number of other examples which show that humans' representations of space are distorted with respect to metric properties (see Tversky, 1992, for an overview).

2.4.2 Landmark, Route, and Survey Knowledge

According to Siegel and White (1975), humans acquire spatial knowledge in distinct stages when their knowledge is elaborated due to interaction with an environment. In each of these stages, a different type of spatial knowledge is acquired. The corresponding types of knowledge are: (1) landmark knowledge, (2) route knowledge, and (3) survey (or configurational) knowledge (see Figure 2.6). The progression from one stage to the other is made by further integrating spatial knowledge (Golledge, Dougherty, & Bell, 1995; Montello, 1998; Ishikawa & Montello, 2006). For example, survey knowledge is the most integrated type of spatial knowledge and can be pictured as knowledge about routes connecting landmarks in a common representation (sharing a common, absolute reference frame; Ishikawa & Montello, 2006).

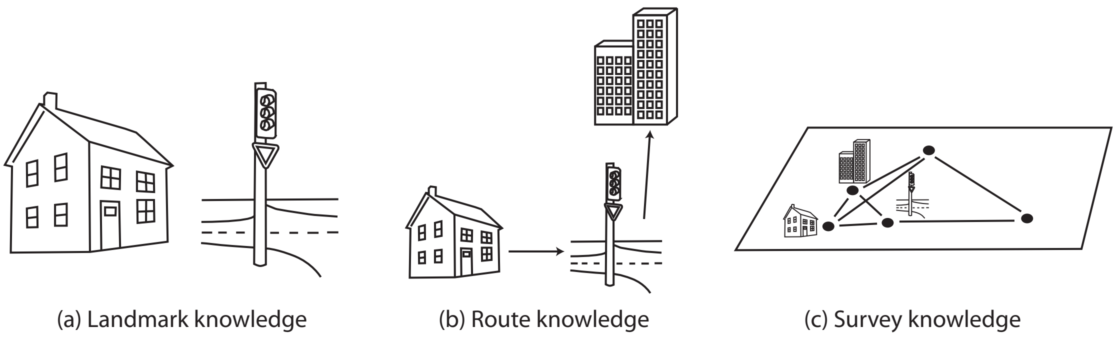


Figure 2.6: Different types of knowledge of environmental space (this picture is a variant of a picture from May, 1992).

Landmark knowledge is the knowledge of discrete objects or scenes. Similarly, Siegel and White (1975, p. 23) define landmarks as "…unique configurations of perceptual events (patterns)". Examples for landmarks are large buildings, traffic lights, or other objects that do not change their location, such as trees. Which real-world objects function as landmark depends on how the environment is structured: In a rural area, usually, there is a smaller density of buildings than in an urban area. Consequently, it is much easier to use a building as a landmark than in an urban area because the building is a distinguishing feature of the environment. In a city, an ordinary building is unlikely to be salient, except from when it has distinguishing features (these can be socially, e.g., if the building is important for everyday life, or perceptual, e.g., if the building is visually—or acoustically or olfactory—easily perceptible). Landmarks can be used as distant orientation points to maintain course—for example, when humans use a salient building such as the Eiffel tower in Paris to maintain their direction. Furthermore, landmarks can designate the beginning and end of route segments and, hence, indicate decision points. For example, if humans know that they have to turn left at the first gas station on the way from their flat to the university, they make use of landmark knowledge to determine my route. In the following, I use the term *'potential landmark'* for objects or configurations of objects on maps that can possibly be used as landmarks during navigation.[12]

[12]I use this terminology because in the case of map use, there is no way to know whether a map object

The second type of knowledge is route knowledge. Routes are interconnected route segments that form a connection between two landmarks. Consequently, route knowledge is knowledge about chains of landmarks and the route segments that connect them (Montello, 1998; Golledge, 1999). In other words, route knowledge can be characterized as knowledge about how to get from a specific point A to a specific point B in the environment (Werner, Krieg-Brückner, Mallot, Schweizer, & Freksa, 1997). When spatial knowledge is available only in the form of route knowledge, route corrections (e.g., because the route is not passable and replanning is necessary) are not possible. Figure 2.6 (b) visualizes route knowledge. In the figure, the buildings and the traffic lights represent landmarks that are remembered and the arrows between them represent knowledge of the connections between them. If humans have acquired route knowledge, their spatial representation can be pictured as knowledge of landmarks (which are point-like conceptualized) that are connected by paths (which are conceptualized as lines) (Siegel & White, 1975). Humans who have only acquired route knowledge (and no survey knowledge) of an environment are able to navigate along the route as long as the order of landmarks along the route corresponds to their knowledge. Furthermore, they are able to detect errors when the sequence of landmarks along the route does not correspond to their knowledge (Siegel & White, 1975).

The third type of knowledge is survey knowledge. Survey knowledge is the most integrated type of spatial knowledge. This type of knowledge enables to make autonomous spatial decisions—a spatial task that is important for visually impaired people (Golledge, 1993). Information about spatial relations between the landmarks is contained in the representation. That is, when humans have acquired survey knowledge of an environment, they are able to determine the (coarse) direction between each of the landmarks they know about. A survey-knowledge representation is organized in an absolute (non-egocentric) reference system, so that directions and distances between places in the environment can be estimated (Klatzky, Loomis, Golledge, & Cicinelli, 1990; Kitchin & Blades, 2002). The reference to landmarks in the environment establishes this absolute reference system (Pick & Lockman, 1981; Newcombe & Huttenlocher, 2003). Often, this type of knowledge is described as bird's-eye-view-like knowledge of an environment (e.g., Thorndyke & Hayes-Roth, 1982). Survey knowledge allows to infer spatial relationships between real-world objects, even though the relationship between the objects has not been directly experienced (Fletcher, 1980). In navigation, survey knowledge is needed for route (re)planning and for taking shortcuts. Due to the high degree of integration, detailed survey knowledge attenuates the risk of getting lost, especially for a visually impaired traveler.

As discussed in Section 2.4.1, survey-knowledge representations are called 'spatial mental models'. The representation of survey knowledge is not necessarily correct in terms of metric distances. Rather, it is subject of systematic distortions stemming—for example, from a hierarchical representation. Yet, coarse grained relations between the objects are represented.

---

that can potentially serve as a landmark is actually used or represented as such in humans' spatial mental model. Consequently, the term 'potential landmark' is more accurate than the term 'landmark' to refer to landmark-type objects (such as buildings) on a map.

Originally, the three types of knowledge have been seen as distinct stages in the learning process when people interact with an environment (Siegel & White, 1975). According to this framework, humans' spatial knowledge develops in a sequential manner: Humans firstly remember landmarks, then proceed to the next stage and acquire route knowledge (without distance information). Finally, knowledge is integrated to a survey representation. Especially the strict sequentially of the knowledge acquisition process has been challenged (Montello, 1998; Kitchin & Blades, 2002; Newcombe & Huttenlocher, 2003; Ishikawa & Montello, 2006). Yet, the terms 'route knowledge' and 'survey knowledge' are widely used to refer to these types of spatial knowledge, often without the implication of a sequential acquisition (e.g., Klatzky et al., 1990; MacEachren, 1992; Kitchin & Blades, 2002; Lobben, 2004; Münzer & Hölscher, 2011; Wen, Ishikawa, & Sato, 2011).

2.4.3 The Content of Representations of Cities

In addition to the different types of spatial knowledge, the content of spatial mental models has been investigated. According to Lynch's (1960) seminal work, humans' spatial knowledge of a city consists of paths, edges, districts, nodes, and landmarks. As detailed in Table 2.3, paths are structures that allow for locomotion (such as walkways). Edges are linear objects that are not used as paths (such as walls). Districts are relatively large city areas (such as Manhattan in New York). Nodes (e.g., intersections) and landmarks (e.g. buildings) often serve as start and endpoints of traveling. The categorization has been empirically tested and proven to be reliable (Aragones & Arredondo, 1985). The content of spatial mental models reflects how humans' structure spatial information.

To decide which information to utter, an interaction system that uses natural-language generation techniques needs to be equipped with knowledge about the content of humans' spatial mental models. Therefore, in the knowledge base in the VAVETaM system, knowledge is structured in categories similar to those suggested by Lynch. Table 2.3 introduces the categories of objects used in the VAVETaM system and the corresponding categories in Lynch's terminology.[13]

As discussed in Chapter 4, the knowledge base is organized using an order-sorted logic. The sorts of the objects basically correspond to Lynch's classification. Yet, some terminological differences exist. As discussed in Lohmann et al. (2011), the term *'track'* is used, instead of the term 'path' (the term 'track' is used by Tschander, Schmidtke, Eschenbach, Habel, & Kulik, 2003, as well; Werner et al., 1997, use the term 'passage' with a similar meaning). The use of the term *'potential landmark'* instead of 'landmark' was discussed earlier (Section 2.4.2). The term *'track configuration'* is used instead of the term 'node' for entities such as intersections of tracks. There are no entities that correspond to districts in the maps that were used in experiments that were conducted (Chapter 3) and in the implementation of a prototype for the language-generation part of the VAVETaM system. However, the region sort is appropriate for such entities. The

[13] Presently, no objects that correspond to the category 'edge' are contained in the knowledge base of the VAVETaM system. However, this is only due to the fact that such objects are not included in the virtual tactile maps used in the empirical evaluation of VAVETaMs.

Table 2.3: The content of spatial mental models of cities (according to Lynch, 1960) and the corresponding categories in this thesis.

Name	**Description (according to Lynch, 1960)**	**Corresponding Category in This Work**
Path	Paths are the channels that potentially enable the observer's locomotion. Examples for paths are streets, trails, walkways, or tunnels	Track
Edge	Edges are linear elements that are not used or considered as paths by the observer, such as shores or walls	
Districts	Districts are medium-to-large sections of the city. Other spatial entities can be 'inside of' a district. Districts are used to an individually varying degree to reason about the spatial layout of a city.	Region (the categories correspond roughly; the region category also includes parks and squares, which the district category does not)
Nodes	Nodes are strategic spots in a city that observers can enter, and which are the intensive foci to which and from which they are traveling. Example nodes are crossings or other convergences of paths.	Track configuration (the categories correspond roughly as Lynch includes larger entities such as squares in in the node category; in contrast, the track configuration sort only refers to configurations of line-like-conceptualized objects)
Landmark	Landmarks are, like nodes, point-like conceptualized. Compared to nodes, the observer does not enter them, they are external. For example, buildings, trees, or mountains can be used as landmarks. Some landmarks are distant ones that can be seen from many angles and distances (e.g., the Eiffel tower in Paris can potentially serve as distant landmark). Usually, the spatial mental model of human observers of a city contains a lot of different landmarks.	Potential landmark (when in the environment or on a map), Landmark (when part of a spatial mental model)

conceptual sorts used in the knowledge base of the VAVETaM system and the system in general are discussed in more detail in Chapter 4.

2.4.4 Spatial Mental Models of Visually Impaired People

Vision is often seen as the most important spatial sense (e.g., Foulke, 1982; see also Chapter 1). With respect to humans' navigation, vision has important functions. In particular, vision allows for an immediate and precise recognition of distant objects (Ungar, 2000). Hence, vision allows for the perception of distant landmarks and makes information about the relations between objects available (Kitchin & Blades, 2002). Therefore, vision facilitates the establishment of absolute reference systems. Without vision, the integration of spatial knowledge that is sequentially acquired (see Chapter 1) is challenging (e.g., Bigelow, 1996). Yet, being able to travel independently is an important part of life. Golledge (1993) states that apart from reading and writing, the most severe problem for visually impaired people are the limitations to travel independently. The quality of human life generally depends on the ability to autonomously make spatial decisions, no matter whether an individual is visually impaired or not. Spatial cognition of visually impaired, especially of congenitally blind people has been the subject of discussion and research (for overviews see Andrews, 1983; Golledge, 1993; Jacobson & Kitchin, 1997; Kitchin & Blades, 2002).

Three theories concerning the ability of visually impaired people to represent space were suggested. They range from denying the ability of congenitally blind people to form a spatial mental model at all to stating that in principle, visually impaired and sighted people perform equally efficient. According to Fletcher (1980), none of the theories necessarily denies the ability of congenitally blind people to acquire (sequential) route knowledge (in contrast to integrated survey knowledge). The following overview follows the discussions by Fletcher (1980) and Kitchin and Blades (2002).

An early theory is the deficiency theory. This theory suggests that the spatial performance of visually impaired (especially congenitally blind) people is generally deficient compared to the performance of sighted people. According to this theory, vision is a necessary prerequisite to form adequate and efficient spatial representations. Visually impaired (especially congenitally blind) people are unable to develop an efficient spatial representation (see also Golledge, 1993, for a discussion). Consequently, following the deficiency theory, the performance of *all* congenitally blind people in spatial tasks is lower than that of sighted people, because they are unable to acquire survey knowledge at all (Fletcher, 1980). The abilities of blind people to learn about and navigate in new environments are clear evidence against this view (as discussed in Section 2.5.2).

The inefficiency theory is based on the assumption that although visually impaired people potentially are able to reach spatial performance of sighted people, they have less efficient ways and strategies of encoding and applying spatial information. The theory implies that visually impaired people make use of the same cognitive processes as people with vision. Visually impaired people are able to build up spatial mental models, but due to inappropriate strategies, their internal spatial representations are less accurate. Humans' cognitive strategies are less well tailored for non-visual perception. Consequently, sighted people outperform visually impaired people in spatial tasks irrespective of how much experience visually impaired people have with the environment.

According to the difference theory, potentially, the spatial mental models of visually impaired people are as accurate and detailed as those of sighted people. This theory states that visually impaired people make use of different strategies and/or their spatial mental models are differently structured than that of sighted people. If there is a delay in ontogenetic development (i.e., the development of individuals in their lifespan), it is possibly confounded with other factors such as general experience with the environment. The theory implies that in principle, visually impaired people are able to perform as well as sighted people in spatial tasks, given appropriate means to acquire spatial knowledge and sufficient experience. Yet, it is possible that the capacity to develop efficient spatial mental models develops later in the ontogenesis of a s person than in ontogenesis of a sighted person (Fletcher, 1980). If the difference theory is valid, visually impaired people are able to perform as efficient as sighted people in spatial tasks if they have enough experience with the environment and suitable means to acquire spatial knowledge.

Evidence clearly speaks against the deficiency theory (e.g., Fletcher, 1980; Landau, Gleitman, & Spelke, 1981; Millar, 1994; Klatzky, Golledge, Loomis, & Cicinelli, 1995; Golledge, Klatzky, & Loomis, 1996; Giudice, Bakdash, & Legge, 2007). Concerning inefficiency and difference theory, Kitchin and Blades (2002) discuss that it is difficult to assess experimentally which of the two is valid because for any result that shows that visually impaired people perform less well than sighted people, it can be argued that this inferior performance stems from lack of experience either with the testing environment or in ontogenesis. Regardless of whether the inefficiency theory or the difference theory are true, it can be concluded that visually impaired people are able to form spatial mental models of their environments if they have suitable sources of information.

2.5 Acquisition of Spatial Knowledge

In the last sections, the discussion focussed on spatial mental models, which are the result of a process (see Downs & Stea, 1977, for a discussion on the process character of spatial cognition): To acquire spatial mental models, humans must acquire and integrate spatial knowledge. As discussed, different external means (e.g., verbal directions, visual and tactile maps) that are based on different sensory (vision, haptics, audition, . . .) and representational modalities (natural language, maps, . . .) or their combinations can be used to acquire spatial knowledge. In the following, I discuss the acquisition of spatial knowledge using different sources in more detail.

2.5.1 Primary and Secondary Acquisition of Spatial Knowledge

Acquiring knowledge about an environment by direct exposure to it or by using external means such as visual or tactile maps or natural-language descriptions are different tasks, which are associated with different cognitive processes (as discussed by Lobben (2004) for direct exposure and visual maps). Spatial knowledge acquisition that is based on direct exposure to the environment is called 'primary acquisition'. For instance, when humans move in an environment they live in, they acquire spatial knowledge by primary acquisition. The other mode of spatial knowledge acquisition is called 'secondary acquisition' (Kitchin & Blades, 2002).[14] Secondary acquisition is based on external representations of space such as visual or tactile maps or descriptions.

As Golledge (1993) points out, spatial knowledge acquisition can be active or passive. For example, when humans traverse an environment by walking or bicycling, they actively interact with it. In contrast, humans perceive environments passively in occasions in which they perceive the environment, but do not make spatial decisions—for example, as passengers in a car. Except for small-scale spaces that are suitable for walking, visually impaired individuals interact passively with their environment (Golledge, 1993). Golledge questions whether passively acquired knowledge can be integrated into a coherent spatial mental model.

Concerning how spatial knowledge is acquired by sighted people, research covers how humans acquire spatial knowledge by primary acquisition (i.e., which cognitive processes and strategies they make use of). In this research direction, various researchers investigated the processes of primary learning of sighted individuals. As discussed in Section 2.4.2, Siegel and White (1975) suggest that primary spatial learning is governed by sequential shifts from landmark to route to survey knowledge (the validity of this theory is questioned, see Section 2.4.2).

The strategies used in map reading, a form of secondary learning, by sighted people have been investigated, as well. MacEachren (1992) conducted a study testing whether different modes of map learning lead to differences in humans' spatial representations. The modes that were compared are: (1) learning a map in an holistic fashion (i.e., showing the map completely), (2) showing only the tracks ('routes', in the original publication) first without showing the rest, (3) showing potential landmarks first, or (4) showing regions (parts of the map) first. The experiment showed that participants who learned the map in the condition where the tracks were shown first learned the map the quickest.

Another research direction concentrates on whether knowledge acquisition from different sources results in similar internal representations; for instance, the result of acquiring spatial knowledge from a map (secondary learning) is compared to acquiring knowledge by direct exposure to an environment (primary learning). G. W. Evans and Pezdek (1980) conducted a series of experiments with sighted participants. In these experiments the spatial mental models resulting from map use were compared with spatial mental models that result from primary learning. Participants were tested using a recognition task. They were asked to choose whether cards showed the configuration of three buildings of a campus were correct or not. Reaction times were measured. Participants were split in two groups: One (student) group that had learned the campus by primary learning prior to the study and one group that had no direct experience and learned the campus using a map. Evans and Pezdek's experiments showed that response times of participants that learned the environment using a map were dependent on the degree of rotation of the

[14] Kitchin and Blades (2002) use the terms 'primary learning' and 'secondary learning'.

recognition cards. In contrast, rotation of the recognition cards did not have a significant effect on the response times of the participants who learned the campus by direct experience. This finding shows that how a spatial mental model is acquired can affect its structure. Spatial mental models from direct experience differ from those that result from map use.

A similar finding was made by Thorndyke and Hayes-Roth (1982). They compared how a building was learned by direct experience and by using a map. In the experiment, sighted people participated. The results show that map learners acquire a survey-knowledge representation. In contrast, knowledge acquisition by primary learning results in route-knowledge—at least, when experience with the environment is moderate. Participants of the map learning group outperformed those of the direct-experience group with moderate experience in tasks that asked for survey knowledge (air-line-distance estimation and estimating the location of a destination room on a map-like representation). The differences between the groups in the survey-knowledge tasks decreased with increasing experience with the environment. With prolonged experience with the environment, sighted people are able to reorganize their procedural route knowledge into a form of survey knowledge.

Taylor and Tversky (1992b) conducted an experiment with which they showed that humans can form accurate spatial representation of an environment when they receive just a linguistic description. Participants read descriptions of the environment. The descriptions were either organized as a virtual tour through the environment (this kind of descriptions is called 'route descriptions' in the publication) or as overview descriptions (this kind of descriptions is called 'survey descriptions' in the publication). The route descriptions made use of an egocentric frame of reference. In contrast, the survey descriptions made use of an absolute frame of reference using cardinal directions. Participants were able to form an accurate spatial mental model of the environment in all conditions (map learning, route description, and survey description).

Further evidence for the functional equivalence (i.e., highly similar performance on spatial tasks; Struiksma, 2011) of spatial mental models acquired by language and by visual perception comes from research that is based on a mental-scanning paradigm (c.f. Loomis, Klatzky, Avraamides, Lippa, & Golledge, 2007; Quinlan & Dyson, 2008). Mental scanning paradigms are based on the idea of an (at least partly) analog internal representation of knowledge (see Section 2.2). If the representation is analog, then the time needed to shift the attention from one feature to another is dependent on the distance between the features (Kosslyn, 1973). According to this theory, it takes longer to answer questions about two features spatially distant than to answer questions about two features spatially close.

Denis and Cocude (1989) conducted a study investigating spatial mental models acquired from different sources. In this study, participants learned a simple environment by secondary learning either from a verbal description or from a visual map. Both groups were similar in that they showed a latency-distance correlation (participants were told to use mental imagery), which indicates that the spatial mental image has the properties of an analog representation. Thus, at least for simple environments, humans are capable of producing spatial mental models on the basis of spatial language that are functionally equivalent to spatial mental models generated by using a visual representation of the environment.

The finding is in line with the findings of Denis and Zimmer (1992), who also discuss empirical evidence that spatial mental models can be generated from verbal descriptions and can show analog properties. Yet, experiments by Avraamides, Loomis, Klatzky, and Golledge (2004) demonstrate that if participants are not encouraged to form a spatial image, performance can differ between visual learning and language-based acquisition. However, if participants are encouraged to form a spatial image, the results of Avraamides and colleagues support earlier findings. The results show that generally, spatial knowledge resulting from language and visual learning can, but must not be equivalent.

In conclusion, humans acquire spatial knowledge either by direct experience of the environment (primary acquisition) or by using external means (secondary acquisition) such as visual maps, tactile maps, or descriptions. Knowledge acquired from a map has a survey-knowledge organization; that is, learning a map leads to integrated spatial knowledge that allows for spatial inferencing (Thorndyke & Hayes-Roth, 1982). Different empirical results show that, in principle, spatial representations derived from language can be functionally equivalent to those that are derived from vision. Yet, they must not be in all cases (Avraamides et al., 2004). All studies agree on the capability of natural language to function as means for spatial knowledge acquisition. This encourages further investigation of combining natural language and tactile maps to create maps for non-visual knowledge acquisition.

2.5.2 Non-Visual Acquisition of Spatial Knowledge

Especially for a visually impaired traveler the acquisition of spatial knowledge is crucial as navigation is a problem, but safe locomotion is, in most environments, not. A skilled visually impaired traveller is able to move securely in the environment. However, orientation (on which correct navigation relies) in the environment is problematic due to limited access to adequate knowledge sources (Leonard & Newman, 1970). The literature reviewed in this section discusses whether visually impaired people are able to build up spatial knowledge by primary and by secondary acquisition. I begin with a discussion of studies that are concerned with how effective visually impaired people acquire spatial knowledge from one source (e.g., tactile maps or by direct experience). Subsequently, studies that compare different ways to acquire spatial knowledge are discussed.

Empirical evidence shows that visually impaired people can acquire spatial knowledge from tactile maps. The general abilities of visually impaired people to read tactile maps were addressed by Landau (1986) and, in another study, by Bérla (1981). Landau (1986) discusses an experiment that showed that a blind (4-year old) child comprehended the fundamental principles of map use and was able perform map tasks using a simple map (but cf. Liben, 1991). Blind adults are mostly able to read tactile maps, yet, training can effect their performance (Bérla, 1981).

Bentzen (1972) tested six participants on a campus location they were unfamiliar with for how well they are able to use a tactile map. All participants were able to plan the route using a map. Subsequently, Bentzen tested whether participants are able to travel the route they had planed. All participants were able to reach the goals on the route (still, some needed help from the experimenter). The participants were asked to make suggestions for improving the map. Three of six participants suggested that the map should include a verbal description of some aspects of the map, which is a further indicator for the importance of developing VAVETaMs.

The findings of Herman, Herman, and Chatman (1983) support that visually impaired people are able to acquire spatial mental models from external haptic representations. Herman and colleagues tested whether using a haptic scale model (i.e., a downsized model of the environment) helps blind people to acquire spatial knowledge. Participants were able to acquire a survey representation of the environment.

In line with the evidence discussed above, Andrews (1983) tested various tactile maps with visually impaired people. The experiments show that visually impaired people are able to use tactile maps of different scale to acquire spatial knowledge.

Another experiment was conducted by Jacobson (1992). Three congenitally blind participants were asked to create a sketch map of a campus location familiar to them using a raised-line drawing board (see Section 3.6, for a discussion of such a drawing board). In a second step, they explored a tactile map and subsequently, they were asked to sketch the map of the campus again. All participants were able to sketch a map of the campus. The sketch maps that were produced after learning the tactile map included more detail and were more similar to the original layout. This indicates that tactile maps can help blind people to understand an environment.

A study by Passini and Proulx shows that visually impaired people are able to acquire spatial mental models of indoor environments by primary acquisition. Firstly, Passini and Proulx (1988) conducted an experiment in which they asked blind and sighted participants to learn a route through a building by primary acquisition. Passini and Proulx conclude that wayfinding tasks are more difficult for blind people, because they have to make and remember more decisions to solve them. Concerning wayfinding efficiency, sighted participants made less errors and more sighted participants were able to complete the route without error. Yet, the number of participants that were able to build a correct model of the route did not differ to a large degree. Passini and Proulx conclude from the results that blind people are able to build up efficient internal spatial representations.

In a similar subsequent study, Passini, Proulx, and Rainville (1990) tested groups of congenitally blind, adventitiously blind, visually impaired, blindfolded sighted, and sighted participants. The experiment showed that "any spatio-cognitive deficiency theory must be rejected" (Passini et al., 1990, p. 115). Yet, there were significant differences between the groups. Passini and colleagues argue that this difference may be attributed to perceptual effects and not to cognitive abilities. In fact, in all tests, non-blindfolded sighted participants performed best followed by congenitally blind participants. In contrast, the blindfolded sighted participants performed worst. The results indicate the abilities of visually impaired people to acquire and apply spatial knowledge.

Several studies have investigated the effectiveness of different ways to acquire spatial knowledge for blind and visually impaired people in comparison. A study by Brambring and Weber (1981) tested whether using a tactile map is more effective than a route description or instructed environmental learning. Participants did not learn a single route, but the whole environment. Then, they were asked to walk along a predetermined route. Participants learned the environment the fastest when they used tactile maps. No significant difference was found between learning the environment by direct experience and learning the environment from the description. Accordingly, the group that learned the environment using a tactile map performed best in terms of navigational performance. The test was repeated six weeks after the initial testing. Participants did not learn the environment again, but relied on the knowledge they had acquired in the initial testing. In the second test, the differences between the groups were even stronger. About 12 weeks after learning the environment, a third test was conducted. Comparable results were obtained. Altogether, Brambring and Weber's results indicate that using a tactile map is the most effective way to learn about an environment for blind people but verbal descriptions can also be used. This indicates the potential of combining these sources of spatial knowledge in a representationally multi-modal interface.

Another study that compared effectiveness of different sources of spatial knowledge was conducted by Espinosa, Ungar, Ochaita, Blades, and Spencer (1998). In two experiments, they compared primary and secondary knowledge-acquisition methods for introducing visually impaired people to an environment. In the first experiment, a route in an outdoor scenario was learned by participants. The participants learned the route either just by direct experience, by a combination of direct experience and a tactile map, or by a combination of direct experience and a verbal description. The group of participants that learned the route in a combination of direct experience and a tactile map showed the best navigational performance. In addition, this group outperformed the other groups with respect to survey knowledge. In a second experiment, Espinosa and colleagues tested the effectiveness of primary learning (without additional cues) and using a tactile map (without primary learning). Between the two learning groups no significant difference was found with respect to navigational performance and with respect to survey knowledge. The results by Espinosa and colleagues indicate that tactile maps can be effective means to acquire spatial knowledge for visually impaired people.

Giudice, Betty, and Loomis (2011) compared whether learning of visual maps and tactile maps results in functionally similar spatial knowledge. In the experiments, participants learned simple routes. Sighted participants performed comparably regardless of the sensory modality in which map-learning was performed. This result indicates functional equivalence of the resulting representations. Furthermore, blind participants who learned the tactile map performed similarly to sighted participants. This shows that, given access to suitable information, blind people have the capability to form spatial representations from external sources such as tactile maps.

Altogether, the empirical evidence shows that tactile maps are usable sources of spatial information for visually impaired people. Furthermore, given access to adequate sources of spatial knowledge, visually impaired people perform similar to sighted people in spatial tasks.

It has also been investigated whether situated verbal descriptions can facilitate non-visual primary acquisition of knowledge about indoor layouts: Giudice et al. (2007) tested whether blindfolded sighted participants benefit from verbal information when acquiring spatial knowledge by direct acquisition of a hallway. Giudice and colleagues compared three different modes of situated verbal descriptions: (1) a mode in which only the user's local position is described (local verbal mode), (2) additional description of distance and geometry for adjacent intersections (maplet verbal mode), and (3) additional information about the overall geometric structure of the layout (global verbal mode). Moreover, in a control condition, participants were not blindfolded when learning the environment. The results indicate that the use of verbal descriptions during free exploration of environments leads to accurate environmental learning. In addition, the result contrast the traditional model that extensive route knowledge is acquired before survey knowledge is developed. Interestingly, there was no significant effect of the type of description in the testing phase. This indicates that the additional distance and global information given in the maplet verbal mode and global verbal mode is not helpful for the spatial knowledge acquisition process. The similarity between visual control group and blindfolded learners with verbal descriptions shows the effectiveness of verbal descriptions to facilitate spatial knowledge acquisition and is additional support for the idea of VAVETaMs.

Giudice (2004, 2006) reports a repetition[15] of the study with visually impaired participants. The performance of this group was comparable to the performance of the sighted group. This result is additional evidence for the effectiveness of verbal descriptions for spatial learning and for the ability of (congenitally) blind people to form adequate spatial mental models. In particular, the results are additional evidence that blind people are not only able to acquire spatial knowledge in the form of route knowledge, but also in the form of integrated survey knowledge, given adequate means for knowledge acquisition.

It has also been shown that such verbal descriptions without locomotion in the environment, but updated according to virtual movements in a virtual environment allow for the non-visual acquisition of spatial knowledge about structured indoor environments, both, for visually impaired and sighted people (Giudice, 2004; Giudice, Bakdash, Legge, & Roy, 2010).

In conclusion, empirical evidence shows that visually impaired people are able to acquire integrated spatial knowledge and to represent this knowledge in spatial mental models, given adequate means for knowledge acquisition. Basically, all studies agree that uni-modal (haptics only) or multi-modal tactile maps are either as efficient as primary learning (i.e., learning by interacting with the environment) (Espinosa et al., 1998) or more efficient than other modes of knowledge acquisition (e.g., Brambring & Weber, 1981; Lahav & Mioduser, 2008b). The effectiveness of tactile maps is evident, even when performance of route traveling is assessed; that is, when route knowledge is important and not survey knowledge (Espinosa et al., 1998). Situated and dynamic spatial descriptions are helpful for primary acquisition in an indoor scenario for blindfolded sighted

[15]The experiment with sighted participants is also reported by Giudice (2004). The experiment with visually impaired people is, as reported, a repetition of the study reported later by Giudice et al. (2007).

(Giudice et al., 2007) and visually impaired people (Giudice, 2004). All studies provide evidence for the capability of visually impaired people to acquire spatial mental models given appropriate means. The empirical evidence discussed suggests that the combination of natural language with (virtual) tactile maps is a promising approach to facilitate the acquisition of survey knowledge for visually impaired people.

2.6 Chapter Summary

In this chapter, I discussed what representations are and which properties they have. Representations are something that stands for something else. While internal representations are humans' mental representations, external representations are physical means that serve as representations.

Representations can be analog (such as maps with respect to the relative location of objects) or symbolic (such as natural language). Many representations are combinations of analog and symbolic representations. For example, in typical road maps, textual labels stand for the names of objects. Such representations are hybrid representations. Maps and natural languages are different representational systems. Representations that combine representational systems are representationally multi-modal.

Humans organize their spatial knowledge as landmark knowledge, route knowledge, and survey knowledge. I call integrated survey knowledge representations of space 'spatial mental models' (as proposed by Tversky, 1993). The ability of visually impaired, especially of congenitally blind people to form such integrated representations of environments has often been questioned. However, empirical evidence shows that visually impaired people are able to perform on a similar level as sighted people in spatial tasks, given adequate means to acquire knowledge.

Empirical evidence shows that physical tactile maps are effective means to acquire spatial knowledge. In addition, studies show that audio-tactile maps, which combine sounds or speech (to present the names of objects) and a haptic presentation are useful. Few controlled experiments exist that investigate the effectiveness of audio-tactile maps. Yet, all empirical evidence points towards the prospects that audio-tactile map systems have as substitutes for visual maps for visually impaired people. Industrially produced and commercially available haptic devices such as the Phantom Omni allow for the haptic presentation of virtual tactile maps.

Empirical investigations with sighted participants show that spatial mental models can be built up from natural-language descriptions of environments. Spatial mental models acquired from descriptions of simple environments are functionally comparable to spatial mental models acquired from maps via vision. Furthermore, spatial mental models that are the result of reading a tactile map are very similar to those that result from learning a visual map. Further evidence for the effectiveness of verbal descriptions is reported by Giudice (2004) and Giudice et al. (2007). They successfully used brief verbal descriptions to augment knowledge acquisition by direct experience of the environment. The approach improved knowledge acquisition of both blindfolded sighted and visually impaired participants. For visually impaired people, knowledge acquisition from tactile or audio-tactile maps is as effective as direct experience (Espinosa et al., 1998) or even more effective than direct experience (e.g., Brambring & Weber, 1981; Lahav & Mioduser, 2008b).

Due to the problems of using textual information in tactile maps, audio-tactile approaches are a promising research direction. Up to date, all the audio-tactile map systems make use of sounds or canned text that is strictly associated to objects or regions on the map and started once the corresponding area or region is touched (or clicked upon). As discussed, verbal descriptions of environments can be used to communicate spatial knowledge. Therefore, it is promising to combine assisting utterances and tactile maps. Possibly, the drawbacks stemming from the sequentiality of perception in tactile maps can be reduced by including verbal utterances that provide additional information to the map user. Aside from some literature that briefly mentions that descriptions could be used (e.g., Petrie et al., 1996), the existing audio-tactile maps do not make use of spatial language.

Taken together, providing assisting utterances in the form of situated assistance for (virtual) tactile map explorations can potentially facilitate knowledge acquisition of tactile maps. Up to now, the effectiveness of the combination of verbal assistance and tactile maps has not been investigated.

Chapter 3

Evaluating Verbally Assisted Maps

3.1 Introduction and Rationale[1]

Maps are important external representations of spatial environments, as discussed in Chapter 1. Currently, access to suitable external representations of survey knowledge is very limited for visually impaired people. While access to appropriate external representations of knowledge is very important for the autonomy of visually impaired people in general, suitable survey representations of space are especially important. Without vision, distant landmarks in the environment, which are important cues for the integration of spatial knowledge to coherent survey knowledge, are not perceivable. Therefore, acquiring spatial mental models by non-visual primary acquisition is cumbersome.

As the review of past research in Chapter 2 indicates, both (virtual) tactile maps and spatial language can be successfully used for the non-visual acquisition of spatial mental models. Yet, as discussed in Chapter 1, tactile maps have drawbacks in terms of speed and accuracy compared to visual maps. It is not possible to describe large and unstructured environments using natural language alone. The combination of spatial natural language and (virtual) tactile maps is a promising approach towards effective non-visual interfaces to spatial information. Yet, such a combination has not yet been investigated. In this thesis, VAVETaMs are proposed as interfaces to spatial information. VAVETaMs provide situated assisting utterances to inform the user about the identity of map objects in the haptic focus and descriptions of the local surroundings on the map. They extend existing audio-tactile map approaches, in which, mostly, the map users are verbally only informed about the names of objects.

I call the type of verbal assistance in VAVETaMs *'extended assistance'*. In this chapter, I present an empirical evaluation of the use of extended assistance for spatial knowledge acquisition from virtual tactile maps. A set of potentially helpful utterances was developed based on a corpus study. In this study, humans were asked to verbally assist a blindfolded reader of a virtual tactile map. Particularly, the analysis of the corpus was focused on the semantic content of the utterances the assistants produced. Additionally, it is of interest how reference to objects is established by human assistants and which reference systems are used. The corpus study is discussed in Section 3.2.

[1]Some parts of this chapter have been previously published (Lohmann et al., 2011; Lohmann, 2011; Lohmann & Habel, 2012).

To test the hypothesis that augmenting virtual tactile maps with extended assistance facilitates knowledge acquisition, a series of experiments were conducted. In order to test effectiveness of knowledge acquisition from a map, adequate testing methods have to be applied. Humans' spatial mental models cannot be directly inspected. Therefore, the application of suitable testing method is important to gain valid results. In order to provide a theoretical basis for the development and application of suitable testing methods of spatial knowledge, I review the relevant literature in Section 3.3.

Three experiments were conducted to test the hypothesis that extended assistance facilitates non-visual tactile-map reading. These experiments were performed as Wizard-of-Oz-like experiments (see Dahlback, Jonsson, & Ahrenberg, 1993, for a discussion of Wizard-of-Oz experiments).[2] The assisting utterances were controlled by the experimenter.

Experiment 1 was performed with blindfolded sighted participants. This experiment had three purposes: (1) it tested the experimental paradigm; (2) it showed whether all testing methods used support the same finding and hence, can be considered valid methods; and (3) it provided empirical evidence whether extended assistance facilitates knowledge acquisition of tactile maps. Experiment 1 is discussed in Section 3.4. With the data of Experiment 1, it was possible to show that the dependent variables are not only valid on the level of convergent validity (on an aggregate level, all tests provide the same result) but on the level of individual data points as well. In Section 3.5, an evaluation method that shows the agreement on the level of individual data points is discussed. Apart from providing results concerning the integration of haptic and natural language representations by sighted participants, one goal of the experiment reported was to test and refine the experimental paradigm before testing visually impaired people.

The potential user group for VAVETaMs are visually impaired people. Giudice (2004) reports on an experiment focused on different forms of verbal descriptions for learning an environment by direct experience. The experiment shows that performance of sighted people is similar to that of visually impaired people. Yet, previous research has not addressed whether visually impaired people perform similar to sighted people when acquiring knowledge from representationally multi-modal external representations. Hence, it cannot be evidently concluded from experiments with sighted participants that extended assistance will have the same effect on visually impaired individuals. For example, visually impaired people may be interested in different aspects of spatial knowledge than sighted people because they need different information during travel. Furthermore, they might have trouble creating a spatial mental model from the external representation (though, empirical evidence reviewed in Section 2.5 does not suggest this). The experimental paradigm and the validity of the testing methods were successfully used in Experiment 1. Subsequently, in *Experiment 2*, Experiment 1 was repeated with a small number of visually impaired people. Experiment 2 is discussed in Section 3.6.

[2] It was avoided to explicitly claim that the utterance generation was performed by the computer system (because it was not necessary). Therefore, I refer to our experiments as 'Wizard-of-Oz-like experiments' instead of 'Wizard-of-Oz experiments'.

The results of Experiment 1 support our hypothesis only in part. Sighted people benefit from extended assistance with respect to their knowledge of landmarks, but not with respect to their knowledge of the track structure. In contrast, Experiment 2 fully supports our hypothesis. Visually impaired people benefit from extended assistance with respect to their knowledge about potential landmarks and about the track structure. In Section 3.7, I discuss potential explanations for the difference between the groups. To test the validity of these explanations, *Experiment 3*, an additional experiment, was conducted. However, the data collected in Experiment 3 does not fully clarify of the cause of the difference in performance between the groups of participants.

Participants were asked to report their exploration strategy after Experiments 2 and 3. Map-exploration strategies were considered to have an effect on map-reading performance. Additionally, it is possible that visually impaired people use different strategies than sighted people do. Based on the reports, two different map exploration strategies were identified. The two strategies are used by both, visually impaired and sighted people. A significant effect on performance is not supported by the data collected.

Altogether, the experiments show that both, sighted and visually impaired people, are able to understand VAVETaMs. Effectiveness of spatial knowledge acquisition of visually impaired people is significantly increased in a condition with extended assistance compared to a condition with less assistance (Experiment 2). This shows that the acquisition of knowledge is facilitated by extended assistance. Both, sighted and visually impaired participants evaluate the assisting utterances given as helpful and complete. Hence, the results encourage the development of the VAVETaM system including the set of assisting utterances as tested in the experiments.

3.2 Basic Message Classes

An important part of developing a language-generation system is determining the information to be expressed (Reiter, Sripada, & Robertson, 2003) and the way it is expressed. Different methods can be used to solve this task—for example, think-aloud protocols, machine learning, or corpus analysis (see Carstensen et al., 2009, for a discussion of the use of corpora in computational linguistics).

The way in which spatial information is linguistically expressed can affect the facility with which people process verbal information. For example, the way sentences are grammatically encoded affects how fast they are understood (Huttenlocher & Strauss, 1968). Regarding descriptions of environments, Taylor and Tversky (1992b) found that route texts are studied longer than survey texts, which suggests that route information is integrated more slowly into memory. A language-based multi-modal map (such as VAVETaM) should respect human preferences for linguistic encoding in order to be effective in the time-constrained situated assistance scenario. Therefore, two questions need to be considered:

- Which information should be given verbally?

- How should this information be given? This question covers the following subquestions: Which reference system should be used? Should information be given in a qualitative or a quantitative manner? How should be referred to the objects on the map?

The information that is conveyed by natural language in VAVETaMs is inspired by the information that human assistants would give in a comparable situation. An interdependence between the map explorer's hand movements and the assistant's utterances occurs: While the utterances are produced with respect to the map explorer's hand movements, the hand movements are affected by what the assistant says.

As an empirical basis for extended assistance, a corpus study was conducted. Assisting utterances produced by humans in the role of assistants that helped a blind-folded map explorer were collected. In this study, the map explorer perceived the virtual tactile map using the Sensable Phantom Omni device (see Section 2.3.1).

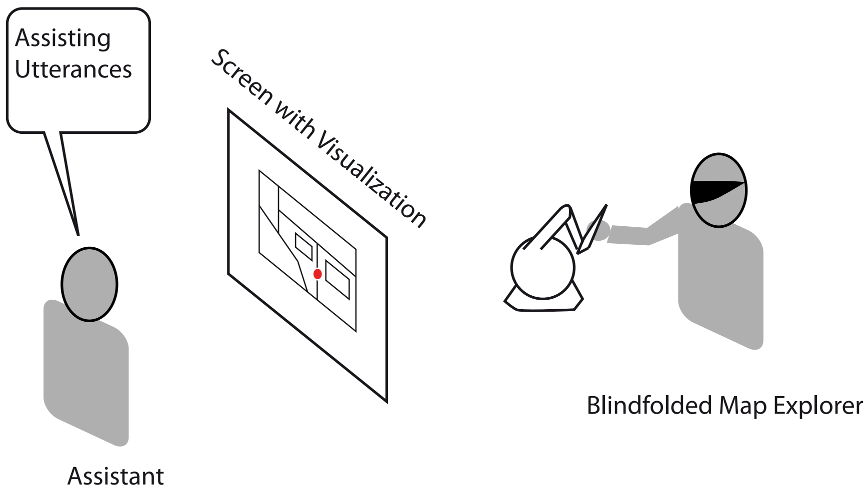


Figure 3.1: The setting in which the corpus was collected.

The assistants saw a visualization of the interface point (the point at which the interaction with the virtual tactile map occurs) in form of a dot moving on a visualization of the map. The movement of the dot corresponded to the map explorer's hand movements and visualized the haptic focus for the assistant (refer to Section 2.3.1, for a discussion of the term 'haptic focus'). The maps had a frame to prevent the map explorer from unintentionally leaving the map. On the maps, streets, buildings, trees, and named squares (e.g., 'Bahnhofsplatz') were present. As Figure 3.1 shows, the assistant saw the map in a vertical orientation on a computer screen (a horizontal display was not possible due to the technical framework). The virtual tactile maps were presented horizontally orientated, parallel to the surface of the table. In preliminary tests, this was identified as the best orientation for enabling map exploration, while preventing the map explorer from quick fatigue. Both participants, the assistant and the map explorer, were instructed that only the assistant should talk; that is, that the co-operative action should not be performed on the basis of a dialogue.

There are no domain experts that could be asked to assist a map user in the corpus study.[3] Assisting a map user verbally to read virtual tactile maps is a novel task, in which people are not usually trained. Hence, five assistants with differing familiarity with the research project were took part in the corpus collection. With different participant/assistant combinations, altogether, 14 map explorations were recorded and evaluated. There were no obvious differences between the assisting utterances produced by assistants that were just briefly introduced to the project and assistants that were familiar with the project, except from some first-time assistants needing some time to train until their assisting utterances were fluent.

The utterances of the assisting humans were recorded and manually transcribed. In a subsequent step, longer utterances were split into utterances that contain only one propositional fact as informational content (see Reiter & Dale, 2000, for a discussion of this technique, and Denis, 1997, for the application of a similar approach). For example, the utterance 'Das ist die Neue Gasse, die rechts in einer Sackgasse endet und links ebenfalls in einer Sackgasse endet' ['This is Neue Gasse, which forms a dead end to the right and forms a dead end to the left as well'] was split into the utterances 'Das ist die Neue Gasse' ['This is Neue Gasse'], 'Die Neue Gasse endet rechts in einer Sackgasse' ['Neue Gasse forms a dead end to the right'], and 'Die Neue Gasse endet links in einer Sackgasse' ['Neue Gasse forms a dead end to the left']. Overall, 1011 utterances were analyzed and, if adequate, split into several utterances. This process resulted in 1229 utterances after splitting. Comments and obviously suboptimal utterances where removed (e.g., 'Hmm, geht das schnell' ['Hmm, this is fast']). Appendix A is an excerpt of the utterances collected (categorized as discussed subsequently, including original and split utterances).

Table 3.1: Message classes for all objects.

Class No.	Informational Content of Utterances
1.1	*Identification* of the object (e.g., using names or referring expressions).
1.2	Utterances expressing the *spatial relation* between objects (e.g., 'left of', 'between').
1.3	Utterances expressing information about *qualitative distance* between objects ('nearby' or 'next to').

Frequently occurring informational categories were determined, subsequently called 'message classes'. A message class is defined by the informational content of an utter-

[3]As confirmed by a professional mobility trainer and a number of visually impaired people in a group discussion conducted at Blinden- und Sehbehindertenverein Hamburg e.V., tactile maps are not commonly used in mobility training of visually impaired people.

ance. Thus, the message classes discussed in the following give an overview of typical information that a human assistant verbally expresses to support a user to explore virtual tactile maps non-visually. They provide a basis for an inventory of assisting utterances for VAVETaMs. Message classes that are relevant for all map objects are shown in Table 3.1. Utterances of the classes 1.1 to 1.3, as introduced in the table, were frequently uttered for every kind of object that was touched including the frame of the map.[4] Table 3.2, discusses message classes 2.1 to 2.3, which contain information given for track objects.

Table 3.2: Message classes for track objects.

Class No.	**Informational Content of Utterances**
2.1	Information about the *extension* of the track; that is, information about what determines the ends of it.
2.2	Information about *junctions* between tracks.
2.3	Information about *geometric relations*, such as parallelism of tracks.

Table 3.3: Message classes for track and potential landmarks objects (not suitable for all objects of these sorts).

Class No.	**Informational Content of Utterances**
3.1	Information about the *location on the map*, expressed in sentences such as 'The town hall is in the upper part of the map.'
3.2	Topological *containment* relations with regions represented on the map, expressed in sentences such as 'The town hall is located on the city square.'
3.3	Information about *extreme positions* on the map ('leftmost', 'rightmost', ...).

Sometimes when track objects and potential landmarks for which giving the respective information was suitable were explored, utterances of the classes 3.1 to 3.3 (Table 3.3) were given. Squares with names were present in the maps used for the collection of the

[4]My description of utterance class 1.1 here differs from the one provided in Lohmann et al. (2011). On the maps used for the collection of the corpus, all objects had names. Therefore, all references to objects were made in a definite manner (allowing to identify the object unambiguously). However, this must not be the case for all map objects. For the purpose of a clearer presentation, I changed the name of message class 1.1 from 'definite identification' to 'identification'. Furthermore, message class 3.3 as discussed here was message class 1.4 in Lohmann et al. (2011). The numbering was changed to make the presentation clearer.

corpus, as well. Consequently, for potential landmarks the additional message class 3.2 was possible and utterances of this class occurred in the corpus.

For regions on the maps, assistants produced some utterances of class 4.1 and 4.2. These message classes are introduced in Table 3.4.

Table 3.4: Message classes for region objects.

Class No.	**Informational Content of Utterances**
4.1	Information about the *extension* or borders of regions.
4.2	Information about the *shape* of regions.

Additionally to the message classes presented, some assistants sporadically gave hints for exploration, see Table 3.5.

Table 3.5: Hints for exploration.

Class No.	**Informational Content of Utterances**
5	*Hints* for exploration such as 'Du hast jetzt alles auf der Karte bis auf die Kirche exploriert' ['Now you have explored everything on the map except for the church'], or 'Da, du warst fast dran, rechts von dir, etwas nördlicher' ['There, you were very close, to your right, a little more to the north'].

The message classes are a specification of the information that is potentially useful to include in the assisting utterances of VAVETaMs. Aside from the question which information to include in the extended assistance, it was attempted to answer the question how the information should be formulated regarding how reference to map objects is made and how spatial relations are encoded.

All assistances were given in a qualitative manner rather than using metric terms (such as stating that a street is 10 cm long). As the message classes discussed above indicate, human assistants inform map users about relative positions between objects or qualitative distances (such as stating that something is nearby an other object).

Human assistants generally refer to the objects on the map with the names of the real-world objects (they were not instructed to do so). They generally do not distinguish between real-world objects, which are represented by geometric objects on the map, and the geometric objects themselves. An example for a real-world object is the 'university' which is represented by a square geometric object on the map. All assistances were formulated in terms of the real-world objects, the intended objects of the verbalization.[5] This is in line with earlier findings on verbalizations of maps (Taylor & Tversky, 1996).

Concerning the reference frames used, most speakers used an absolute reference system (for a discussion of reference systems, see Section 2.3). Some used a mixture between cardinal directions ('north', 'south', ...) and projective terms ('left', 'right', ...). Research has indicated that the preference for reference systems is dependent on the task that speakers have to solve. Hund, Haney, and Seanor (2008) asked participants to give directions in two different conditions. In one condition, participants were told that they would give directions to a person actually driving through the town, in the other condition, they were told that they would give directions to a person looking at a map. Participants used significantly more cardinal directions when they solved the task to describe a route to a person looking at a map. As the human assistants that participated in our study were told that they should describe the map to a blindfolded map explorer, this finding is in line with the finding of Hund and colleagues. In addition, most map users do not explore the virtual variants of tactile maps with steady hand movements, but rather move back and forth when they focus on a map object. This makes the application of an reference system based on the direction of the hand movements of the explorer impossible. The movement direction changes too quickly to verbalize assisting utterances using a movement-based reference system.

In Section 2.3.2, I discussed that suitable reference systems have to be used when maps are described. In our study, the terms 'north' and 'south' were frequently used to refer to the directions 'up' and 'down' on the assistant's screen and, correspondingly, away from and towards the body center of the map explorer (as described, the assistant's map visualization was oriented vertically and the virtual tactile map horizontally) with a combination of a fixed use of 'left' and 'right' instead of 'west' and 'east'. Preliminary tests showed that while using cardinal directions for the north-south axis is understood without problems, some people have problems understanding cardinal terms for east-west axis. However, in the experiments reported in Sections 3.4, 3.6, and 3.7 the projective terms 'up' and 'down' were used instead of 'north' and 'south'. Using cardinal terms would have restricted VAVETaMs to maps in which the top-direction and the north direction fall together (which is not always the case, especially for city maps).

With the corpus study described, we developed an idea about potentially helpful information to be included in assisting utterances for VAVETaMs and about how this information should be formulated. Whether respective assisting utterances facilitate knowledge acquisition was tested in controlled experiments reported in the following sections.

[5] In contrast, humans make reference to both, geometric objects and intended real-world objects when they verbalize sketch maps while they are created (Tappe & Habel, 1998). The data of Tappe and Habel suggest that whether verbalizers know the purpose of a drawing has an effect on the choice if objects are referred to as real-world objects or geometric objects. All our assistants knew that they helped the map explorer to understand the map. In addition, next to the map objects, the real-world names of objects were presented as written labels on the screen.

3.3 Testing Map Effectiveness

The corpus study discussed enabled the development of a set of assisting utterances that likely facilitate knowledge acquisition of tactile maps. Yet, it is not ensured that verbal assistance as proposed in fact improves the effectiveness of the knowledge acquisition process. In the following sections I report an empirical evaluation whether extended assistance as discussed is more effective than verbally just informing map users about the names of objects (which can be achieved with existing systems). To test the effectiveness of VAVETaMs with extended assistance, an experimental setting was developed. Three experiments were conducted. They are discussed in Section 3.4, 3.6, and 3.7. As the goal of using VAVETaMs is effective acquisition of spatial knowledge, valid and reliable methods to test spatial knowledge have to be applied in empirical evaluations of their effectiveness. The corresponding literature is reviewed in Section 3.3.1. In Section 3.3.2, I discuss the type of maps used in the experiments. In Section 3.3.3, I provide a rationale for the selection of participants.

3.3.1 Testing Methods for Human Spatial Knowledge

Overview of Methods

Suitable testing methods for spatial knowledge have to be applied to test whether VAVETaMs are effective means to acquire spatial knowledge. The methods to investigate spatial knowledge discussed in the literature range from asking participants to make direction and distance estimates to sketch mapping (the methods are discussed later in this Section). All methods—in particular, sketch mapping—have been controversially discussed with respect to their validity and reliability. A valid method measures exactly the construct that it is supposed to measure; for example, a valid thermometer measures temperature (and not, e.g., density). Reliability of testing methods refers to repeatability of the measurements (e.g., Blades, 1990). When a thermometer is used to measure the temperature of cooking water, if the thermometer is reliable, the result stays the same when measured at different time points (given that the pressure stays constant).

All tests of spatial cognition that are based on observing participants' performance in a task, such as distance estimations or sketch mapping, are based on representations of humans' knowledge. Participants produce externalizations of spatial knowledge, such as the sketch map or estimation data. Externalizations of spatial knowledge are external representations of humans' internal representation. As such, they are *re*-representations and the result of an extraction (Golledge, 1976) or a readout process (Hegarty, Montello, Richardson, Ishikawa, & Lovelace, 2006) of the internal representation of spatial knowledge.

The readout or extraction process can have an effect on the outcome. For example, in the experimental setups discussed in Sections 3.4, 3.6, and 3.7, participants acquire knowledge of VAVETaMs. They build up a more or less accurate internal representation of the map. After doing so, they are asked to sketch the map. When they sketch the map, they have no access to the original map. Participants base their sketches on their internal representations. They produce sketches that are as similar as possible to the original map as they remember it. However, when the map was not well-learned, they might have acquired incomplete knowledge. For instance, they might not remember the relative location of two buildings, but know that two buildings were on the map. Asking them to sketch a map forces a decision whether to leave out one building or to randomly place it with respect to the other building. If participants leave out one building, all data based on analyzing the sketch map leads to the wrong conclusion that they did not acquire knowledge about the building at all. If they include the building at the correct position, the data leads to the wrong conclusion that they acquired knowledge about the location of the building. In contrast, if the building is included at the wrong position, the data leads to the conclusion that participants internally represent the location of the building wrongly, which is not correct either.

Similar examples can be constructed for all tasks that are based on observing spatial knowledge on the basis of an external representation. While all psychological experiments have to deal with variance in the data, the major problem with respect to research that is based on re-representations is that it is unclear whether the variance is systematic (i.e., with respect to the example discussed, whether most participants include the building or leave it out). Due to systematic and unsystematic variance, the re-representations produced do not necessarily correspond to the spatial mental model they are based on in a one-to-one fashion. Especially due to the possibility of systematic but uncontrolled variance, the validity of different methods has been discussed in the literature.

Different taxonomies of methods for the study of spatial cognition have been discussed in search for valid and reliable methods to test different aspects of spatial cognition (e.g., Montello, 1991; May, 1992; Kitchin & Blades, 2002). Methods can be classified as follows (based on May, 1992, and Kitchin & Blades, 2002):

- Estimation-based methods
- Reconstruction-based methods
- Chronometric methods
- Recognition-based methods
- Qualitative methods

When *estimation-based methods* are applied, participants are asked to estimate either distances between two objects in the spatial layout or angles between two objects (for examples see Curiel & Radvansky, 2004; Stamm, Altinsoy, & Merchel, 2011). There are two subclasses of distance-based methods: methods with absolute quantitative judgements and methods with relative quantitative judgements. In methods with absolute judgements, participants answer questions such as ‘How far is A away from B?’ In methods with relative judgements, participants are given a reference distance or a reference angle—for example, with questions such as ‘If A is 100 meters away from B, how large is the distance between C and D?’ A third group of estimation-based tests rely on ordinal judgments. Participants compare objects, which are located in space, either pairwise or by setting up ranks. In general, estimation-based methods can be performed with a number-based outcome (verbal or written, e.g., 'the distance is 100 m') or a spatial outcome (graphical or haptical, e.g., asking participants to sketch the length of a distance in comparison). Using quantitative-estimation based methods is most appropriate when an underlying theory, for instance, on distortions of spatial mental models, is to be tested.

In addition to the discussed quantitative absolute or relative estimation-based methods, a related method, statement verification, was used by researchers to test knowledge about a spatial environment (Taylor & Tversky, 1992b; Brunyé & Taylor, 2008a, 2008b). In this task, participants are asked to indicate whether statements about a spatial environment are true or false.[6] A true/false-estimation approach seems suitable to test the effectiveness of VAVETaMs, because it allows to test whether participants in general acquired knowledge about the layout without being too sensitive to small systematic distortions (as an estimation-based task based on quantitative outcomes would be). Additionally, it is suitable to test for survey knowledge acquired from maps because it does not demand for transformations of spatial knowledge (such as rescaling or changing perspectives). Furthermore, as a verbal task it is suitable to be used in experiments with sighted and visually impaired participants without any adaption, such that the results are comparable.

Reconstruction-based methods rely on the task to physically reproduce a model or a depiction of the original spatial layout. For example, sketch mapping and model building are reconstruction-based methods. Especially sketch mapping is a classic, frequently used method to study spatial cognition. For example, in the classical works of Lynch (1960) and Downs and Stea (1977), sketch maps are used to analyze humans' spatial knowledge. Sketch mapping is a flexible (May, 1992) and easy-to-use method with which most participants are familiar. Yet, it has been controversially discussed whether sketch mapping can be considered a valid method or whether the outcomes are confounded by other abilities, such as general sketching ability (see also Kitchin, 2000, for a discussion).

An alternative to sketch mapping is the use of multidimensional-scaling (MDS) on the basis of estimation data (May, 1992; Kitchin & Blades, 2002). MDS is a method in which a two-dimensional representation is generated from estimation data. MDS algorithms minimize the difference or conflicts between the estimates and construct a two-dimensional, map-like representation of the data (Kitchin & Blades, 2002). Kitchin and Blades question the validity of MDS methods. In contrast to sketch mapping and reconstruction-based methods, in MDS methods, participants cannot change the relationships between objects when necessary. A similar criticism of MDS is presented by Newcombe (1985). Newcombe concludes that the problems of sketch mapping were possibly exaggerated and the benefits of MDS overestimated. In summary, the literature suggests that sketch mapping is more reliable than MDS as a method which results in spatial, two-dimensional data.

[6] Participants are asked to respond to qualitative statements or questions. In this respect, true/false estimations are a qualitative form of assessment. Yet, when multiple data points are collected, by aggregation of the values, a quantitative outcome is generated. Therefore, I categorize this task as a quantitative task.

Montello (1991) discusses the influence factors of different reconstruction-based methods on the data collected. Montello suggests that the need for scale translations (e.g., in sketch mapping) and other spatial transformations explain the possibility that results vary when different methods are applied. In addition, the independence or non-independence between pairs of data is an issue. A measure is non-independent when it depends on one or all of the other measures (Montello, 1991). For example, in sketch mapping, the position of potential landmarks can depend on the position of streets, the courses of streets are dependent on each other, and so on. Consequently, sketch mapping is a non-independent form of data collection. Mapping techniques require participants to simultaneously represent the relative locations of three or more objects or places in an environment. The size of this representation is smaller than the size of the space that is mapped, therefore, participants need to perform a scale translation. When knowledge acquired from maps is tested in sketch mapping, the sketch maps can be of similar size like the maps originally learned. This minimizes the need for scale transformations and suggests that sketch mapping is appropriate to test knowledge acquired from maps. Consequently, according to this reasoning, sketch mapping is a suitable method to test survey knowledge acquired from maps such as VAVETaMs.

Montello compares the strengths and drawbacks of sketch mapping as data-collection method to the strengths and drawbacks of other methods. Besides the drawback of strong non-independence of estimates (as all objects have to be depicted in relation to each other), Montello concludes that mapping is very useful to test the accuracy of survey knowledge (with some restrictions on the conclusions that should be drawn from the data).

In conclusion, the literature indicates that sketch mapping is an appropriate method to test spatial knowledge acquired of VAVETaMs. However, as discussed later in this section, data on the use of sketch mapping (and other reconstruction-based methods) with visually impaired people is sparse.

A few empirical results concerning the validity and reliability of sketch maps exist. Blades (1990) investigated the reliability of sketch maps by asking participants to sketch a familiar route and then draw the same route a week later. If sketch maps are reliable measures of spatial knowledge, the map pairs should be similar. In fact, Blades' study indicates that the evaluation of sketch maps is a reliable method. Billinghurst and Weghorst (1995) studied the validity of sketch maps as measures for spatial mental models of virtual environments by calculating the correlation of the goodness of sketch maps with self-reported feeling of orientation. They conclude that sketch mapping, as a measure of spatial mental models, is most useful in dense worlds containing a large number of objects. To explain and quantify the effect of drawing ability, Bell and Archibald (2011) used a task in which a complex figure had to be reproduced by the participants. The results show that sketching ability in non-geographic tasks influences the sketch-mapping ability. Yet, the data indicate that sketching ability (as indicated by the results from the task to reproduce the complex figure) only accounts for less then 7 % of the variance of the results. In Section 3.5, I discuss the validity of sketch maps based on the data gained in Experiment 1.

Chronometric methods are based on measuring the time participants need for a task—for example, to respond to a question (Kitchin & Blades, 2002). The latency times that participants show for different tasks are measured as outcome variables. For instance, the experiments based on mental scanning paradigms that were mentioned in Section 2.5 are based on chronometric methods (e.g., Denis & Cocude, 1989; Avraamides et al., 2004). Chronometric methods are suitable to test theories about how spatial knowledge is represented. They are less suitable to test whether spatial knowledge is acquired at all, which is a necessary first step in evaluating an interface to spatial information. Hence, they are not suitable to test whether verbal assistance facilitates knowledge acquisition of virtual tactile maps.

In *recognition tasks*, participants are asked to identify a spatial configuration. For example, participants are shown different maps with the task to identify the map that correctly represents the environment they previously learned. G. W. Evans and Pezdek (1980) used a recognition task in which participants had to identify correct configurations of either three states or three buildings among wrong configurations. A similar method is used by Kitchin (1996) and in the experiments that I report later in this chapter.

Some researchers have empirically compared different quantitative methods. The results of these investigations differ with respect to whether different methods support the same conclusions. For example, Howard, Chase, and Rothman (1973) showed that sketch mapping, producing a model, and estimation of distances correlate to a high degree. However, as the study only tested for distances between buildings, the scope is limited. May (1992) compared two estimation-based methods and a reconstruction-based method. While the methods allowed for conclusions which are in agreement for some aspects of spatial knowledge, conclusions for other aspects of spatial knowledge disagree. In another study reported by Kitchin (1996), differences between different methods were found and explained with task demands.

In summary, the literature does not indicate a single quantitative method that can be considered valid and reliable to test spatial knowledge acquired from VAVETaMs. Yet, as discussed, a true-false-estimation approach is likely suitable, especially, as it can be applied for sighted and visually impaired participants without adaption. Furthermore, among the reconstruction-based methods sketch mapping is considered more valid than MDS methods, and recognition tasks are promising approaches.

Aside from quantitative approaches, *qualitative methods* can be used to study human spatial cognition. Examples for qualitative methods that are used in the study of spatial cognition are talk-aloud protocols (e.g., Ungar, Blades, & Spencer, 1997), semi-structured interviews (see also Kitchin, 2000, for a discussion of applications in the study of spatial cognition), or the analysis of descriptions of environments (e.g., Taylor & Tversky, 1992a; Tenbrink, Bergmann, & Konieczny, 2011). Talk-aloud protocols are not useable to evaluate VAVETaMs, because it is not possible to ask participants to verbalize while they receive verbal assistance. In Experiment 2 and 3, we used a qualitative method in the form of interviewing participants about their experience with VAVETaMs and the map-exploration strategies they used in addition to quantitative methods (see Section 3.6 and 3.7).

Methods: Conclusions

As mentioned, no suggestion for a single method to study spatial cognition is made in the literature. Rather, it is suggested to use different methods in combination (Jacobson, 1998a; Kitchin & Blades, 2002). In the experiments reported, this approach was adopted. In the empirical validation of VAVETaMs (1) *configuration questions* (a true-false estimation task), (2) *sketch mapping*, and (3) a *recognition task* were combined.

The choice for a true-false estimation method over an estimation-based method that includes estimations in terms of numeric values was made for the following reasons: Firstly, estimation-based methods in which estimations are made on the basis of numeric values demand for a very fine-grained and metric representation of spatial knowledge. They are suitable if—for example, systematic distortions in well-learned spatial environments are in the focus of research. In contrast, the empirical work reported was centered on the effectiveness of the knowledge acquisition process rather than on the structure of the internal representation. Therefore, the main interest was to test whether verbal assistance facilitates knowledge acquisition at all (even if the spatial mental model acquired is coarse-grained). The metric accuracy of knowledge is tested in the sketch mapping task. In addition, the question method that we applied allowed to include questions about the structure of the configuration of tracks (e.g., whether two tracks intersect) and, therefore, covers an important aspect of spatial knowledge which quantitative estimation-based methods do not cover.

As mentioned, in addition to the configuration-questions task, participants were asked to sketch the map they learned. As a graphic reconstruction-based technique, sketch mapping has different demands than the verbal task discussed. Hence, the combination of methods ensures that the overall results are not confounded by the demands of a single task. As discussed, because the knowledge is learned from maps, no rescaling or other spatial transformations are requested from participants, which indicates appropriateness of the method. Furthermore, the produced sketch map is a spatial representation that can easily be inspected by the researcher—for example, to identify common mistakes in the participants spatial knowledge. As discussed, using sketch maps is considered more valid in the literature than generating a spatial representation using MDS from a number of distance estimates.

Thirdly, a recognition-based task in which participants had to identify parts of the map was used. Again, this task is very different from the other two and assures validity of the experiment because the task-demands differ from the other two tasks.

Testing the Spatial Knowledge of Visually Impaired People

Important for the research reported is how visually impaired people can be tested for their spatial knowledge. Kitchin and Jacobson (1997) provide an overview of methods that can be applied to analyze the spatial mental models of people with visual impairments (I am not aware of newer publications in this direction). Kitchin and Jacobson distinguish between route-based techniques and tests to measure survey knowledge[7]. As VAVETaMs are intended to be efficient means for the acquisition of survey knowledge, methods to measure this type of knowledge are suitable. Among the tests discussed by Kitchin and Jacobson are haptic adaptions of graphic tests, such as sketch mapping, and recognition tests. Kitchin and Jacobson state that sketch mapping can be troublesome for visually impaired individuals due to the need of classification, simplification, symbolization, and transformation to a bird's eye view. Usually, blind individuals do not have experience with these tasks. Yet, Jacobson (1992, 1996) has successfully used sketch mapping in experiments with blind and visually impaired participants. In the experiments, Jacobson used a special raised-line drawing board similar to the one that we used in Experiment 2, reported in Section 3.6. Downs and Stea (1977) report on a blind geography student that was able to produce accurate sketch maps as well. Giudice (2004) uses model-building, another reconstruction-based method, successfully with visually impaired people—yet in the model-building task, no significant result is achieved, in contrast to other tasks conducted.

Concerning recognition-based methods, Kitchin and Jacobson (1997) conclude that recognition tests are likely to have "great utility for measuring the configurational knowledge of individuals who are visually impaired or blind" (p. 367). The authors state additionally that up to date, they are unaware of any study that has utilized this method to asses the configurational knowledge of visually impaired individuals. I am not aware of any study with visually impaired participants that has used a recognition-based method (in Experiment 2, a tactile version of a recognition-based method was used, see Section 3.6). In line with various researchers discussed above, Kitchin and Jacobson suggest the application of mutually supportive tests, as all test have a number of problems and unknown qualities.

Due to the sparse literature on which methods can successfully be applied to test visually impaired people, the tasks that were used to test visually impaired people in Experiment 2 were of comparatively explorative nature. The configuration-question task was applicable without adaption, which allows comparability between the experiments. A tactile drawing board was used for non-visual sketch mapping. Furthermore, a tactile version of the recognition task was created. The visual variants of the tests applied are discussed in detail in Section 3.4.2. The tactile adaptions of the three tests are discussed in Section 3.6.2.

3.3.2 Using Maps in Experiments

When experiments with maps are conducted, which kind of maps is used has to be considered. In various studies, map-like representations that do not resemble usual maps of real-world areas were used (e.g., Denis & Cocude, 1989; Curiel & Radvansky, 1998, 2004; Lawrence, 2011). See Figure 3.2 for an example of a map-like depiction that does

[7]Kitchin and Jacobson use the term 'configurational knowledge'. As mentioned in Chapter 2, the term has the same meaning like the term 'survey knowledge', which is preferred here because it is the more frequently used term. Furthermore, it describes the type of knowledge better because it implies that knowledge is stored in a way that can be described as 'overview-like'.

not resemble real-world maps that was used by Curiel and Radvansky (2004). As the figure shows, this map is not similar to a common map with respect to its structure, the objects shown, and so on. This kind of maps is called ‘pseudo maps’ (Blades, Ungar, & Spencer, 1999). In other studies, maps that represent an actual real-world area were used. I call this kind of maps ‘real-world maps’. For example, Jacobson (1998a) reports on two preliminary studies in which tactile maps of real world areas were used to test whether they help visually impaired individuals. Espinosa and colleagues (1998) used tactile real-world maps of a part of Madrid to investigate whether blind participants can use them to acquire knowledge of the urban environment.

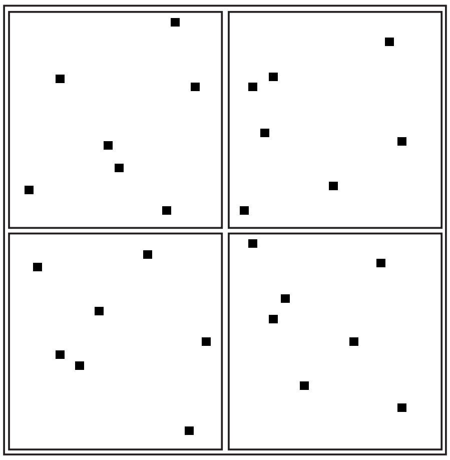


Figure 3.2: The map-like depiction used by Curiel and Radvansky (2004).

Frequently, maps are used that are, with respect to their structure and the objects shown, like real-world maps, but do not depict a real-world area (e.g., Bérla, Butterfield, & Murr, 1976). This kind of maps is called ‘fictional maps’ (Noordzij, Zuidhoek, & Postma, 2006) in the following. Fictional maps are maps that represent non-existent spatial environments that are similar to real spatial environments. Fictional maps resemble real-world maps concerning the type of objects they show and their structure, but they do not depict real-world areas. In other words, fictional maps are maps that plausibly could be used to do the map tasks one usually does with a map, such as acquiring survey knowledge to plan a route or for later navigation, if the spatial layout they depict existed in the real world.

In the experiments reported below, fictional maps were used. As the experiments were performed in a repeated-measures design (that is, each participant is tested repeatedly under different conditions), the use of multiple maps was necessary to avoid carry-over effects (i.e., participants remembering the map from a previous learning condition). Different maps of comparable complexity were needed. To make sure that the maps have the same complexity, it was chosen to use maps that had the same number of streets, crossings, intersections, buildings, and so on. This was only possible with fictional maps.

Additionally, the maps had the same scale and dimensions. Therefore, using real-world maps was not feasible. Furthermore, the use of fictional maps made sure that no participant had prior knowledge of the environments represented.

Aside from the question whether real-world maps, pseudo maps, or fictional maps are used, the complexity of experimental maps is an issue that has to be considered in an experiment with maps and a study-recall procedure. The necessity to remember and recall the maps restricts the complexity that they can have. Possibly, the experimental result is biased if maps are too complex to be remembered. In a preliminary study, visual variants of the maps that were used in the experiment were tested for whether they are easily to remember. Participants had no difficulties to sketch the maps completely and answer a set of questions about them without mistakes.

3.3.3 Visually Impaired and Sighted Participants and Map Experience

Blindfolded sighted participants took part in Experiment 1. The choice for a setting with blindfolded participants rather than with visually impaired participants was made for two reasons. Firstly, reading a map is a complex skill that both, sighted and visually impaired people have to learn (see Liben, 1991; Millar, 1995; and Kitchin & Blades, 2002; but c.f. Landau, 1986). Visually impaired people are not always familiar with maps. The effects of (un)familiarity with map-like representations could not be anticipated. Testing blindfolded sighted participants, who are familiar with maps, ensured homogeneity among the participants. This homogeneity was considered beneficial to test the empirical paradigm and to estimate the training effort needed (yet, in Experiment 2, no problems with respect to inhomogeneity were encountered). Secondly, visually impaired people are a small group. This group is already to a large degree involved in experimental testing in various research contexts. One goal of Experiment 1 was to provide insights how and whether haptic and natural language representations are integrated by sighted participants. A second goal of the experiment was to test and refine the experimental methods and to identify potential problems or improvements with respect to VAVETaMs before testing visually impaired people.

Some studies have compared how visually impaired and sighted people understand spatial language. Both groups perform similarly in a lot of tasks when appropriate testing procedures are used (Klatzky et al., 1995). Yet, differences in performance have been shown concerning the preferred reference systems in spatial language (Noordzij et al., 2006; Struiksma, 2011). Research in the area of non-visual spatial-knowledge acquisition suggests a similar performance (e.g., Giudice, 2004), but is somewhat diverse (see Section 2.5, for a thorough discussion). Altogether, the current state of research did not allow for an anticipation whether visually impaired and sighted participants perform similarly in acquiring knowledge from VAVETaMs. Consequently, Experiment 1 was repeated with visually impaired participants to ensure validity of the results for the potential user group. In this experiment—Experiment 2—a mixed group of congenitally blind, late blind, and visually impaired people took part. The group of participants was mixed because the potential user group of the VAVETaM system is diverse with respect to their visual impairment, as well.

Because problems with respect to the inhomogeneity of visually impaired people's experience with maps were anticipated (but not encountered), participants in Experiment 2 were asked to indicated their experience with maps on a 1–5 scale. The statements that were used to assess their experience were the following:

- Statement 1: 'Please indicate your experience in reading visual maps on a scale from one to five.' ['Bitte schätzen Sie Ihre Erfahrungen im Kartenlesen visueller Karten ein.'] (1 = none [keine], 5 = very experienced [sehr erfahren])
- Statement 2: 'Please indicate your experience in reading tactile maps on a scale from one to five.' [Bitte schätzen Sie Ihre Erfahrungen im Kartenlesen taktiler Karten ein.'] (1 = none, 5 = very experienced)
- Statement 3: 'I use tactile maps regularly.' ['Ich nutze taktile Karten regelmäßig.'] (1 = never [nie], 5 = very frequently [sehr regelmäßig])

Table 3.6: Map experience of visually impaired participants in Experiment 2. (The data of participant 7 was excluded from evaluation, as discussed in Section 3.6).

Participant	**Onset Age of Impairment**	**Answer**		
		Statement 1	**Statement 2**	**Statement 3**
1	Congenitally	1.00	3.00	2.00
2	51 years	5.00	1.00	1.00
3	30 years	2.00	2.00	1.00
4	6 years	3.00	3.00	2.00
5	0.5 years	1.00	5.00	1.00
6	Congenitally	1.00	2.00	1.00
8	Congenitally	1.00	5.00	1.00

For an overview of the answers see Table 3.6. As the table shows, all three congenitally blind participants reported to have no experience with visual maps (statement 1). Of the two participants with an early onset of visual impairment (0.5 years and 6 years), one reported medium experience with visual maps and one none. With respect to experience with tactile maps (statement 2), participants with late onset of visual impairment reported the lowest experience. The participants who were congenitally blind and those with early onset reported, in average, a larger experience. All participants reported to never use tactile maps (statement 3), except one congenitally blind participant and one participant with an early onset who reported to seldom use tactile maps. In contrast to our expectations, all indicators for map experience do not correlate significantly with the results of the configurations questions task or the tactile sketch maps. This indicates that experience with visual or tactile maps is not a predictor for the performance with VAVETaMs for the group of visually impaired people.

In Experiment 3 (a repetition of Experiment 1 with track-only maps), sighted participants were asked whether they are able to orient themselves with maps in cities new to them and whether they sometimes loose orientation in cities new to them even when they use a map. Interestingly, in this experiment, significant correlations with the number of correctly answered configuration questions after learning virtual tactile maps in the extended-assistance condition were found (see Section 3.7.4). This indicates that general map-reading ability predicts sighted people's performance with VAVETaMs (at least for the track-only maps used in Experiment 3, in Experiment 1, participants were not asked to indicate their map-reading abilities).

The results of Experiment 2 show differences in performance between sighted and visually impaired participants. Experiment 3 was performed to explain these differences. This experiment was, as Experiment 1, performed with sighted participants. The rationale for this choice of participants was that the finding of Experiment 1 (the experiment that was performed with sighted participants) contradicted some of the initial expectations. In contrast, the results of Experiment 2 (with visually impaired participants) were in line with the expectations. Hence, Experiment 3, an altered version of Experiment 1, was conducted with sighted participants to gain insights why they perform differently than initially expected and than visually impaired participants did.

It was not possible to control the sampling of the participants with respect to age or level of education. Furthermore, it was not possible to assess underlying abilities with standardized tests: It is unknown which general abilities underlie the understanding of representationally multi-modal information displays such as VAVETaMs. Controlling for verbal intelligence would have been possible, but only would have covered a part of potential underlying abilities. Because the assisting utterances are mostly short and do not contain complicated vocabulary, it was not anticipated that verbal intelligence is a predictor for performance. In addition, assessing for verbal intelligence would not have controlled for a full account of individual differences. Furthermore, at least for navigational performance, it has been shown that the influence of small-scale spatial abilities as assessed in standardized intelligence tests is small (Hegarty et al., 2006). It is unclear to which degree the ability to read and understand maps is affected by this type of spatial abilities.

Therefore, a repeated-measures paradigm was designed, which provides results that are not confounded by most individual differences. Theoretically, it is still possible that differences between the groups confound the results (i.e., that the processing of information in only one of the conditions is affected by another variable than by visual abilities). Yet, there are no theoretical reasons to assume this.

3.4 Experiment 1: Sighted Participants

3.4.1 Introduction

The goal of the research reported is to develop an effective means to acquire survey knowledge non-visually. Assisting utterances in combination with virtual tactile maps are proposed to facilitate knowledge acquisition. The experiment reported in this section has the purpose to evaluate our proposal for the VAVETaM system and to test the empirical setting developed. The experiment was performed in a Wizard-of-Oz-like manner (i.e., the experimenter imitates intelligent behavior of a computer system). We tested two types of verbal assistance for tactile-map explorations, which differ with respect to the information provided. One condition with less verbal assistance is called 'simple-assistance condition'. In this condition, participants received only utterances that inform them about the names of the objects. This type of verbal assistance is similar to the proposals in other discussions of audio-tactile map systems (e.g., Jacobson, 1992; Wang et al., 2009; Buzzi et al., 2011). In the extended-assistance condition, participants additionally received information about map objects that do not carry proper names (such as crossings) and information about relations between map objects (e.g., about buildings that are located close to a street). The extended-assistance condition is inspired by humans' performance in the task to verbally assist a map user. It is based on the set of potentially helpful utterances that was developed on the basis of the corpus study reported in Section 3.2.

The spatial knowledge acquired in the simple-assistance condition is compared to the knowledge acquired in the extended-assistance condition. The hypothesis is that more precise spatial mental models are acquired in the extended-assistance condition compared to the simple-assistance condition: Additional verbal information and haptically perceived information are successfully integrated in the process of knowledge acquisition.

3.4.2 Method

Participants: Blindfolded Sighted Participants

Twenty-four participants initially took part in the experiment. Data of two participants had to be excluded from analysis due to technical problems. One participant was not able to pass the test that concluded the training procedure and the corresponding dataset was therefore also excluded. Data of three additional participants were collected, leading to a total of 24 evaluated participants (14 males, mean age: 24.7 years, *SD*: 3.3 years). All participants were compensated by partial course credit or on a monetary basis. They were naïve about the purpose of the experiment. All participants gave written informed consent and reported to speak German on a native-speaker level. All participants used their self-reported primary hand (data of 17 right-handed participants was evaluated) for the exploration of the virtual tactile map. They had never seen the experimental setup and were not given any feedback on their performance during or after the experiment.

Material and Procedure

Material. Two different virtual tactile maps and corresponding utterances with two different name sets where created to avoid carry-over effects in the repeated-measures design. The utterances were started by the experimenter using a custom-built interface.

Maps. The virtual tactile maps were haptically explored by the participants using a Sensable Phantom Omni interface (see Section 2.3.1) attached to an Apple iMac. The maps were fictional maps of similar complexity regarding the number of intersections,

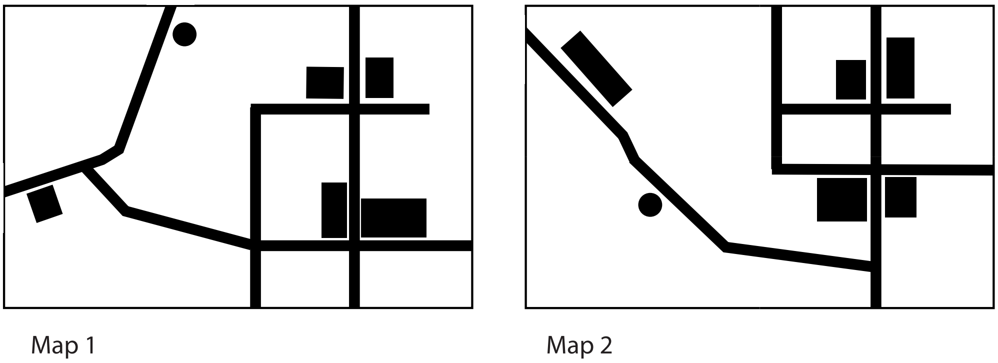


Figure 3.3: Visualization of the two maps used in Experiment 1.

the number of parallel streets, potential landmarks, and dead ends (as discussed in Section 3.3.2). Both maps included five tracks and six potential landmarks (one tree and five buildings). See Figure 3.3 for a visualization of the maps used. A pre-study showed that depictions of the maps could be completely remembered when learned visually. The maps were modeled for haptic interaction using Autodesk's 3D Studio Max. The 3D models were presented with the device using the Sensable OpenHaptics toolkit[8].

Table 3.7: Name sets used in Experiment 1.

	Set 1	**Set 2**
S	Poststraße, Humboldtstraße, Lärchenweg, Goethestraße, Hegelstraße	Hochstraße, Dorfstraße, Amselweg, Blumenstraße, Bergstraße
LM1	Hauptbahnhof [main station], Universität [university], Christuskirche, Bertolt-Brecht-Schule	Rathaus [town hall], Gedächtniskirche, Anne-Frank-Schule, Museum [museum]
LM2	Aldi	Lidl
LM3	Eiche [oak]	Buche [beech]

Assisting Utterances. The assisting utterances were recorded before the study. They were given in German, spoken by a 26-year-old male native German speaker (the author). The participants heard the utterances via headphones. Two name sets (i.e., sets of proper names for the map objects) were created that could be used for both maps (the amount of tracks and potential landmarks is identical on both maps). See Table 3.7 for an overview of the name sets used. The name sets consisted of street names for the

[8]http://www.sensable.com

tracks (S), and of names for potential landmarks. The latter were of the following types: (LM1) names signifying the function of the potential landmark, for example 'Anne-Frank-Schule'; (LM2) proper names such as brand names for chains of stores, for example 'Aldi'; and (LM3) class names, such as 'Buche'.

The utterances in the corpus were classified in message classes based on their informational content (Section 3.2). In the context of the Experiment 1, the identification message class is particularly important. By stating an utterance of the identification message class, the assistant informs the user of the tactile map about the identity of the map object that is explored. Usually, this is done by stating the name of the map object in combination with a demonstrative. If an object does not have a name it can often be identified by referring to objects with names. For example, intersections can be identified by stating the names of the streets that form the intersection (see example (3a) below).

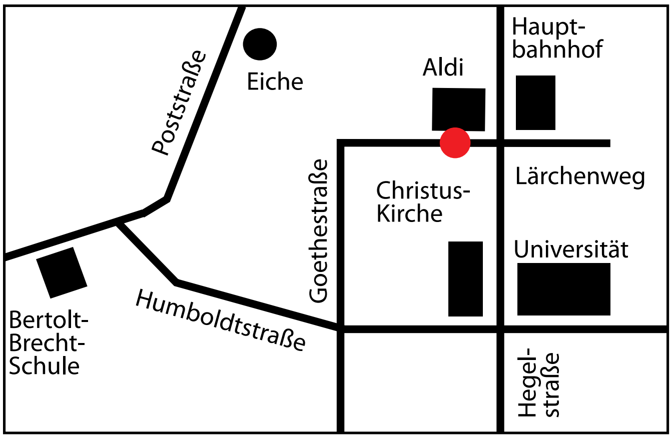


Figure 3.4: Visualization of one of the maps used with one of the name sets. The small dot between the buildings 'Christuskirche' and 'Universität' indicates the participant's exploration position.

In the simple-assistance condition only information about the names of objects in the haptic focus (i.e., the object or part of the object the map explorer wants to gain information about, see Section 2.3) was given. Only utterances of the identification message class for objects with proper names were included. Consequently, no information for map objects without a proper name was given. In the maps used for the study, this affected intersections and dead ends, which were not verbally identified in the simple-assistance condition.

Example (1a) is a translation of the assisting utterance that was given when the track 'Lärchenweg' was explored. This utterance was given in both conditions. In the extended-assistance condition, more assisting utterances were given (see Appendix B.1, for a transcription of the original German utterances for Map 1 and 2 with the first name set): Examples (1b) and (1c) are translations of utterances of the identification message class for the left frame of the map and the building called 'Bertolt-Brecht-Schule'. Note that utterances that use deictic reference are time critical; that is, they should only be given when the participant actually explores the object that is talked about. For example, utterance (1a) should only be given when the participant is currently exploring the corresponding track. Figure 3.4 shows a position on this track with a dot. Considering this position is the participant's exploration position, giving utterances such as (1b) or (1c) would be inappropriate.

(1a) This is Lärchenweg.[9]

(1b) This is the left map frame.

(1c) This is Bertolt-Brecht-Schule.

In addition to utterances such as (1a)–(1c), information that a human assistant would potentially include was given in the extended-assistance condition. (2a)–(3c) are examples of translations of assisting utterances given in this condition for track objects. The assisting utterances (2a)–(2d) are suitable when a participant explores the track 'Lärchenweg'—for example, at the position marked with a dot in Figure 3.4. The extended-assistance condition included verbal information about the extent of tracks (i.e., what determines the end of a track; see example (2b)). Furthermore, the set of utterances included information about the intersections a track has (2d) and information about spatial and geometric relations with other tracks and landmarks (see (2a), (2b), and (2c)). As the identification utterances, they were only given when the participant explored the part of the map that they were about.

(2a) Lärchenweg is parallel to Humboldtstraße.

(2b) Lärchenweg ends to the left in a corner with Goethestraße and to the right in a dead end.

(2c) Above Lärchenweg, there are Aldi and Hauptbahnhof.

(2d) Lärchenweg intersects with Hegelstraße.

Furthermore, for parts of tracks that were close to landmarks or between landmarks, assisting utterances were given that stated this relation (see utterance (3c)). Additionally, intersections and dead ends were identified (see utterances (3a) and (3b)). Like the identification messages, these message use deictic reference and are time critical.

(3a) This is the intersection between Goethestraße and Humboldtstraße.

(3b) This is the dead end that forms the right end of Lärchenweg.

(3c) Now, you are below Aldi.

[9]The purpose of the translations is to illustrate the content for the reader. Determiner conventions are purposefully neglected.

Examples (4a) and (4b) are translations of assisting utterances for a potential-landmark object. For these objects, the set of utterances in the extended-assistance condition included utterances that state the relation to other map objects (see (4a)) and, if appropriate, the global location in the map (see (4b)).

(4a) Bertolt-Brecht-Schule is located below Poststraße.

(4b) Hauptbahnhof is located in the upper part of the map.

Control of Assisting Utterances by the Experimenter. The experiment was performed as Wizard-of-Oz-like experiment. The experimenter controlled the playback of the utterances using custom-built software developed for that purpose. Figure 3.5 shows a visualization of the experimenter's interface for the extended-assistance condition. The squares represent buttons in the software that, when pressed, started the playback of the utterances.

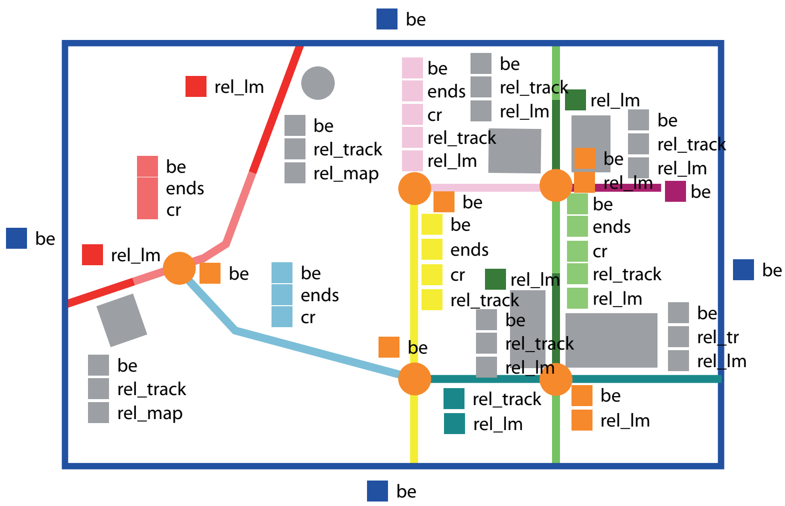


Figure 3.5: The experimenter's interface with color-coded buttons starting the assisting utterances.

The experimenter started assisting utterances when participants explored a map object with their hand movements. When a map object or a part of it was in the haptic focus—that is, when the explorative hand movements indicated that the respective object is the subject of exploration—corresponding utterances informing the map user about the object and its surroundings were started. Color-indication was used to state which button belonged to which map object or to which part of an object. As in the corpus collection, a dot moving corresponding to the participant's exploration movements visualized the haptic focus of the participant for the experimenter. The abbreviations next to the buttons (as shown in Figure 3.5) indicated what class of assisting utterance each button started. The abbreviation be indicated that the button started the playback of an utterance of the identification message class for the object focused upon, as exemplified by the utterances (1a)–(1c). A button labeled with ends started utterances that inform about the extension of the object (see (2b)). cr started utterances informing the participant about the intersections of the street (see (2d)). rel_track started utterances expressing information about spatial and geometric relations between the corresponding object and a track object (see (2a)). rel_lm started the expression of information about the relation between the corresponding map object and potential landmarks (see (2c) and (4a)). A button labeled rel_map started an utterance including information about the position of the corresponding object on the map (see (4b)).

Pre-studies and the corpus study (Section 3.2) indicated that if a map object is explored, giving information of the identification message class should precede any other information. Hence, the experimenter started an utterance of the identification class prior to any other utterance in the extended-assistance condition. The other utterances for that object were given in the order that the experimenter found most appropriate. Unnatural repetitions of utterances were avoided. To ease this, the buttons for utterances that were started once were marked in the interface.

Testing Methods. Following the map exploration with one of the assistance conditions described, the spatial knowledge that participants gained by the exploration of the map was tested. Different, mutually supportive methods were used: (1) asking questions about the configuration of objects on the map, (2) sketch mapping, and (3) a task in which participants needed to recognize visualized parts of the map in a setting similar to a jigsaw-puzzle.

Configuration Questions. As a first task after map exploration, participants answered questions concerning the spatial layout of the map. A similar approach was used previously by Taylor and Tversky (1992b) and Brunyé and Taylor (2008a, 2008b) to test spatial mental models acquired from different types of spatial descriptions. The configuration-question test consisted of 20 questions on spatial relations between objects and on the configuration of the objects. Where it was possible, these questions asked for relations between objects and object configurations that were not explicitly stated in the pre-recorded assisting utterances for the extended-assistance condition. The experimenter posed the questions in a predetermined individual random order.

As reported in the results, a subsequent analysis revealed that the effect of the experimental condition on questions involving potential landmarks and those involving only tracks was different. Therefore, the discussion of questions follows the distinction of questions involving only potential landmarks and those involving tracks, as well. 10 questions involved spatial relations including potential landmarks. For translations of these questions, see Table 3.8.[10] These questions are called 'landmark questions'. 10 questions were about the track configuration. For translations of these questions, see Table 3.9.

[10] The numbering of the questions is used to enable referencing to questions in this description; the questions were, as described, posed in a random order.

Table 3.8: Translation of the questions for Map 1 with the first name set that only involve landmarks. (* Question 5 was wrongly formulated in one set and therefore excluded from the detailed analysis, as described later.)

Question Number	**Translation of the Question**
1	Is Eiche left of Hauptbahnhof?
2	Is Eiche left of Hegelstraße?
3	Is Bertolt-Brecht-Schule left of Christuskirche?
4	Is Hauptbahnhof above Universität?
5*	Is Christuskirche below Aldi?
11	Is Eiche right of Christuskirche?
12	Is Hegelstraße left of Bertolt-Brecht-Schule?
13	Is Bertolt-Brecht-Schule above Eiche?
14	Is Hauptbahnhof below Aldi?
15	Is Hauptbahnhof right of Universität?

Table 3.9: Translation of the questions for Map 1 with the first names set that involve tracks.

Question Number	**Translation of the Question**
6	Do Blumenstraße and Bergstraße form a T-intersection?
7	Are Goethestraße and Hegelstraße parallel?
8	Does Lärchenweg form a dead end to the right?
9	Do Hochstraße and Bergstraße intersect?
10	Do Goethestraße and Lärchenweg meet?
16	Do Bergstraße and Amselweg form a T-intersection?
17	Does Goethestraße form a dead end?
18	Do Poststraße and Lärchenweg form an intersection?
19	Do Amselweg and Blumenstraße intersect?
20	Do Hochstraße and Dorfstraße meet?

These questions are called 'track questions'. For the original German questions see Appendix B.4. The answering options were 'yes', 'do not know', and 'no'. 10 questions were answered correctly with a 'no' and 10 questions with a 'yes'. A correct 'yes' and correct 'no' were evaluated as correct answers. A wrong 'yes', a wrong 'no', and 'do not know' were evaluated as wrong answers.

In the experimental procedure, each name set occurred with each map. Therefore, sets of configuration questions were developed for the each combination of maps and name sets. This resulted in four sets of 20 questions. The questions of these sets were matched with each other. For example, question 1 always asked for the relation of two landmarks that were relatively distant to each other and was always correctly answered with 'yes' and question 17 always asked whether two streets are parallel and was always correctly answered with 'no'. Two sets of questions were created for each map. This corresponds to the use of two different name sets for each map. The individual questions for each name set always asked for a spatially equivalent fact. For example, in the first map for the first name set the first question was 'Is the Eiche left of Hauptbahnhof?' In the second name set the map object that was called 'Eiche' in the first set was called 'Buche'. The map object that was called 'Hauptbahnhof' in the first set was called 'Anne-Frank-Schule' in the second set. Consequently, the first question for the first map with the second set of names was: 'Is Buche left of Anne-Frank-Schule?'

Sketch Task. After answering to the configuration questions, participants were asked to sketch the map on a sheet of paper. For a discussion of the validity and reliability of sketch maps as assessment methods for spatial knowledge see Section 3.3. Section 3.5 discusses the validity of the data generated in this experiment.

The frame defining the dimensions of the map was printed on the paper for sketching. All 48 sketches were evaluated by the author and an independent rater. Both raters were naïve with respect to the condition in which the sketch map was produced and to which sketch maps were created by the same participant. The rating was performed in two respects. These two respects correspond to the two knowledge types identified in the principal component analysis for the configuration questions, described in Section 3.4.3. They were: (1) to what degree does the sketch resemble the original map concerning the course of tracks, their parallelism, and the junctions they have and (2) to what degree are potential landmarks represented at the correct position? The rating was performed on a 5-point Likert-type scale. A rating of 1 is associated with 'does not reflect the original' and 5 is associated with 'reflects the original precisely'.

Recognition Task. The third test performed was a recognition task comparable to a jigsaw puzzle. A visualization of the map that was previously explored by the participants was split into quadrants. Participants were given a set of possible map parts for each quadrant and were asked to decide which one is correct.

The goal positions and orientation of the parts were given. For each position, there were six options: the correct solution and five parts with a wrong spatial layout of a track and/or potential landmark. Each potential part fitted to each potential part of the other quadrants: there were no potential landmarks or tracks on any part that would have led to an inconsistent picture when combined. Refer to Figure 3.6 for an example of the options for the upper-right part of the visualization of the map.

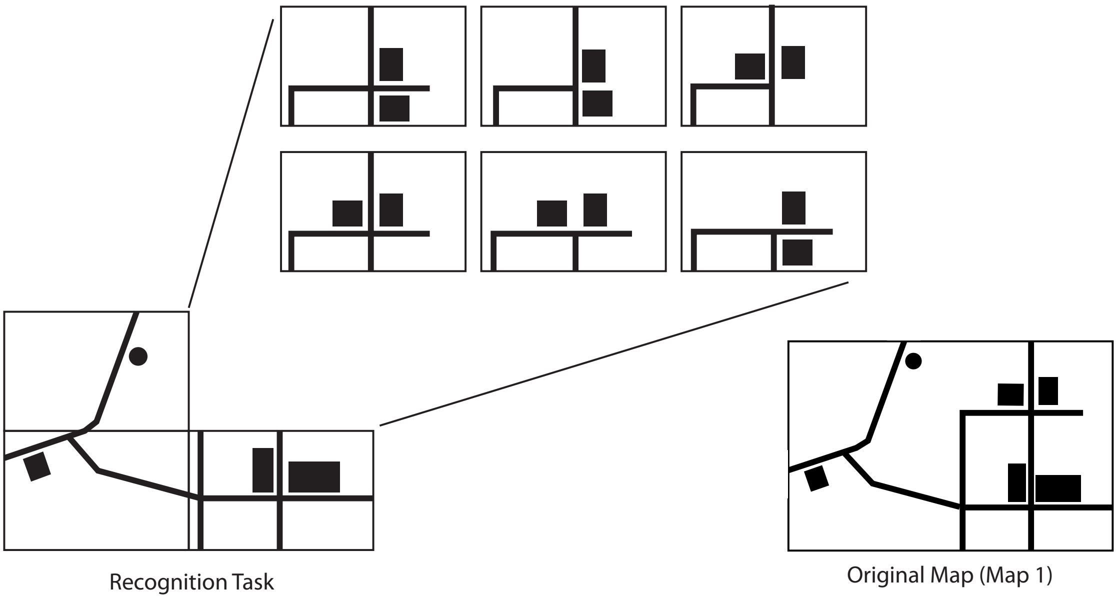


Figure 3.6: Example for the recognition task showing options for the upper-right part of Map 1. Choice options were given for the other three quadrants, as well.

Procedure. The experiment was performed as repeated-measures experiment with the assistance condition as within-subject variable. This design was chosen to overcome the problem of individual differences in spatial abilities that possibly confound results (see Wen et al., 2011, for a recent discussion of individual differences in spatial abilities). To avoid carry-over effects, two different maps and name sets were used for the different experimental conditions. As discussed, to control for order effects confounding the main research question, the experiment was fully counterbalanced concerning the order of assistance conditions, maps, and the name sets. Accordingly, the second learning procedure was performed with the map, the assistance condition, and the name set that were not used in the first condition. Consequently, the extended-assistance condition occurred as often as first learning condition as the simple-assistance condition. Furthermore, the Map 1 and the first name set occurred as often as Map 2 and the second name set. Following this principle, each assistance condition was performed as often with each map and each name set as the other one.

To ensure that participants understood the interaction with the verbally assisted maps, they were trained in different aspects of it. Firstly, they were introduced to the haptic device and explored some standard examples of the Chai 3D[11] haptic toolkit, which is a framework for haptic rendering. Then, they were interactively introduced in virtual

[11] http://www.chai3d.org

tactile maps by exploring a training map. In this part of training, first, they were assisted by the experimenter who gave assisting utterances similar to those provided in the experiment. Subsequently, they received prerecorded utterances controlled by the experimenter (as in the experimental conditions). The map used for training purposes was different from the two maps used in the experimental conditions. In the training procedure, the map objects had artificial names such as 'A-Building' or 'Alpha-Street'. Participants were allowed to see a visualization of the map after haptic exploration to enable a matching of existing concepts to the tactile map scenario. To conclude training, participants were tested for their ability to identify the shape of objects in virtual tactile maps (such as a triangle and a square). This was done because preliminary experiments had shown that the ability to find objects and to identify shapes of objects is an important aspect of reading virtual tactile maps. Furthermore, participants were asked to follow a complex track structure without leaving it while receiving and following assisting utterances given by the experimenter. This task was included because preliminary experiments had indicated that the ability to follow line-like objects is crucial for understanding the track network. The training procedure including the training test took 45–60 minutes.

The complete experiment took approximately two hours including training. After training and a short break, participants read written instructions about the time they had for exploration and the tasks they had to solve after learning. Information about the tasks was included to minimize order effects resulting from knowledge of the tasks in the second condition from solving them in the first condition. For the same purpose, an example map similar to the ones used in the experimental conditions was printed on the instructions. Participants were instructed to learn the map in a way that they would find the route from each potential landmark to each other (see Appendix B.2, for the instruction).

Each map exploration was limited to eight minutes. After map exploration, first the configuration-questions task was performed, then participants created the sketch map, and finally did the recognition task. A ten-minute break was made before the participants performed the second exploration with the same time constraint as the first exploration.

3.4.3 Results

Configuration Questions.

Taking all questions into account, the average result showed that participants were able to answer significantly more questions correctly ($t(23) = 8.08$, $p < .001$) when they learned the map in the extended-assistance condition ($M = 14.04$, $SE = .61$) than in the simple-assistance condition ($M = 8.46$, $SE = .42$) and the effect is large ($r = .86$) (see Cohen, 1992; Field, 2005, for a discussion of effect sizes).[12] Figure 3.7 visualizes the result. Overall, the data does not indicate a strong effect of order: In the condition conducted firstly, participants were able to answer comparable many questions correctly ($M = 11.54$, $SE = .73$) as in the condition conducted secondly ($M = 11.04$, $SE = .83$).

[12] An α-level of .05 was used for all calculations reported in this thesis.

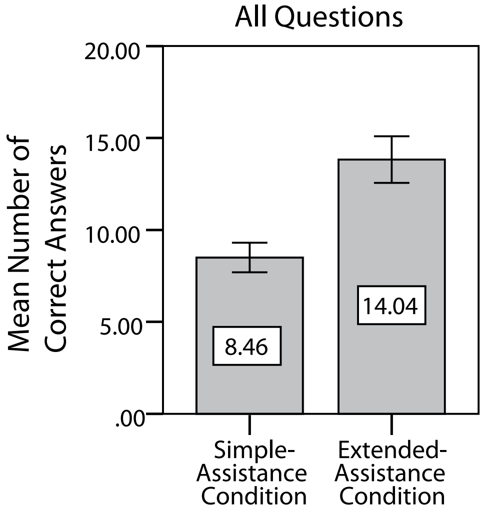


Figure 3.7: Mean number of correct answers to all questions of the configuration questions task. Error bars represent the 95 % confidence interval of the mean.

In the subsequent analysis, question 5 was excluded due to a mistake in formulation of this question in one set of questions. In the analysis reported subsequently, we assumed that which name set and which map was used does not have an effect on the answers on the questions (no significant effects of the map or the name set used were found in the analysis of the remaining 19 questions).

Screening the data, it was obvious that the assistance condition had a strong effect on some but not on all questions. Whereas the answers to the landmark-related questions showed large differences between the conditions, the answers to the questions that only involved track knowledge did not. To support this theory, a principal component analysis was performed. Two components were extracted, corresponding to the two types of questions. Independency of the components was not assumed. Therefore, oblique rotation was chosen as rotation method.

Interestingly, the components correlate negatively (component one correlates with –.202 with component two). The component loadings are shown in Table 3.10. The questions involving potential landmarks load highly on factor one, those that do not involve knowledge of potential landmarks load highly on component two. The components together explain about 40.59 percent of the variance. An explanation why question 8 does not load as expected on component 2, which reflects track knowledge, cannot be given with the current data.[13] As the component loadings support the hypothesis that the questions only about tracks and the questions that involve landmarks measure different components of spatial knowledge, they were analyzed separately.

[13] Question 8 includes the direction of the dead end. It is, in this respect, different from the other track questions. Possibly, this difference affects the outcome. I thank Carola Eschenbach for pointing this out.

Table 3.10: Component loadings of the questions on the two knowledge dimensions.

Question Number	**Component 1 (Landmark Knowledge)**	**Component 2 (Track Knowledge)**
Landmark Questions		
11	.832	.028
2	.784	–.038
1	.776	.125
13	.749	.382
14	.685	.426
15	.684	.247
3	.669	.121
12	.656	–.029
4	.546	.246
Track Questions		
16	–.380	.669
20	–.342	.643
9	–.204	.527
6	–.319	.483
17	–.071	.384
19	–.204	.335
7	–.362	.334
18	–.175	.261
8	.298	.020

Figure 3.8 shows the mean number of correct answers to the two types of questions. As shown, extended assistance leads to a larger number of correctly localized landmarks, the knowledge of the track structure seems not to be affected by the learning condition.

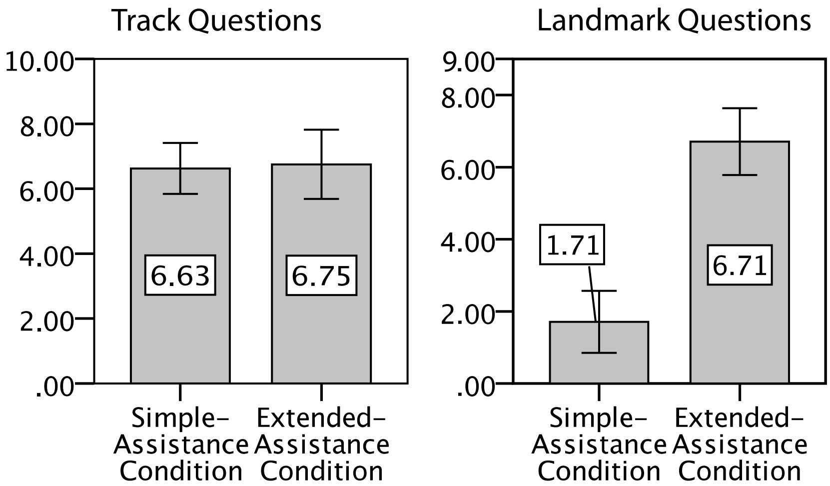


Figure 3.8: Mean number of correct answers to the track questions and the potential-landmark questions in the configuration-questions taks. Error bars represent the 95 % confidence interval of the mean.

Separate mixed-design ANOVAs were calculated for each of the types of questions. In both types, there were no significant main effects of gender and handedness. Likewise, there were no significant interaction effects of these variables with assistance condition. Therefore, these variables were stepwise excluded from the model for further analysis.

Concerning the track questions, there were no significant interaction effects of assistance condition with the order of conditions (simple assistance first or extended assistance first), with the map used (which map was used with simple assistance) or with the set of names (which set of names was used with simple assistance). Therefore, these variables were stepwise excluded from the model, as well. Participants were not able to answer significantly more questions after learning the map in the extended-assistance condition.

A paired-sample t-test of equivalences (Wellek, 2003) was used to control whether the number of correct answers to the track questions can be considered equivalent in both map-learning conditions. In the analysis, we used a liberal symmetrical equivalence interval of .50 and an alpha level of .05 (two tailed), resulting in the following t value: $t = .23$. This value is lower than the critical constant, therefore, the number of answers can be considered statistically equivalent among the conditions. Note that a ceiling effect can be ruled out; the average shows that there is still room for improvements on the scale. With respect to the order of conditions, participants were able to answer slightly less track questions in the first condition they participated in ($M = 6.25$, $SE = .42$) than in the second condition ($M = 7.13$, $SE = .47$) they participated in.

For the landmark questions there was no interaction effect of condition with the set of names (which set of names was used with simple assistance). Hence, this variable was excluded from the model. However, there was an interaction effect of condition with the order of conditions (simple assistance first or extended assistance first; $F(1,21) = 7.41$, $p < .05$, $r = .51$). Specifically, participants were able to answer more landmark questions in the condition they firstly took part in ($M = 4.83$, $SE = .68$) than in the second condition they took part in ($M = 3.58$, $SE = .65$). Furthermore, there was an interaction effect of condition with the map used (which map was used with simple assistance; $F(1,21) = 5.56$, $p < .05$, $r = .46$). In particular, participants were able to answer more questions in the simple assistance condition with Map 2 ($M = 2.67$, $SE = .68$) than with Map 1 ($M = .75$, $SE = .30$). Therefore, these variables were left in the model. With respect to the effect of the assistance condition, the counterbalanced setting controls bias from these variables.

In contrast to the track questions, there was a highly significant and large effect of assistance condition on the number of correct answers for the landmark questions ($F(1,23) = 118.50$, $p < .001$, $r = .92$). Participants were able to answer significantly more questions correctly after having learned the map in the extended assistance condition.

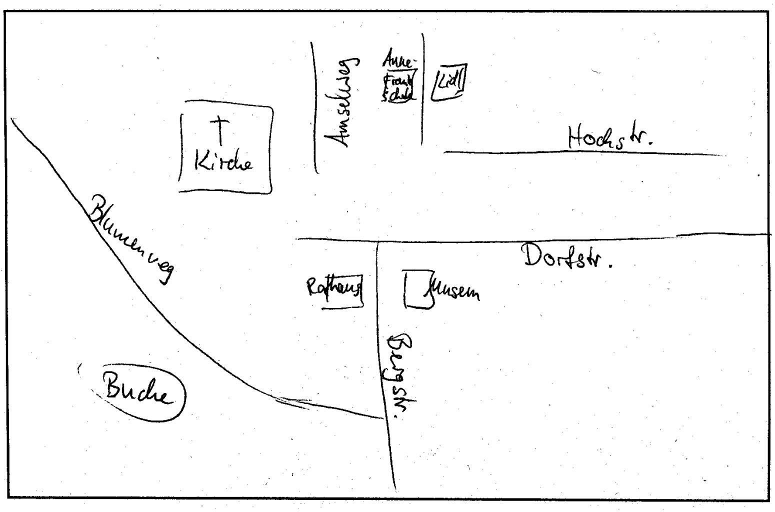


Figure 3.9: Example of a participant's sketch of Map 2 after learning the map in the extended-assistance condition.

Sketch-Mapping Task. Two raters evaluated the sketches. The intraclass-correlation coefficient was calculated to test the agreement between the raters. The agreements of the track ratings (*ICC(3,1)* = .68) and of the potential-landmark ratings (*ICC(3,1)* = .72) were calculated separately (Shrout & Fleiss, 1979). The raters' agreement was fair in both ratings. The values of the raters were averaged for further analysis. Figure 3.9 shows an example of a participant's sketch map. Figure 3.10 shows the mean ratings of the sketches. The values indicate that potential landmarks depicted corresponded better to the map explored when it was learned in the extended-assistance condition.

Separate mixed-design ANOVAs[14] were calculated for the landmark-configuration ratings and the track-structure ratings. In both ratings, there were no significant main effects or interaction effects with assistance condition of gender, handedness, which map was used with which condition, the order of the conditions, or which set of names was used with which condition. Therefore, these variables were stepwise excluded from the model for further analysis.

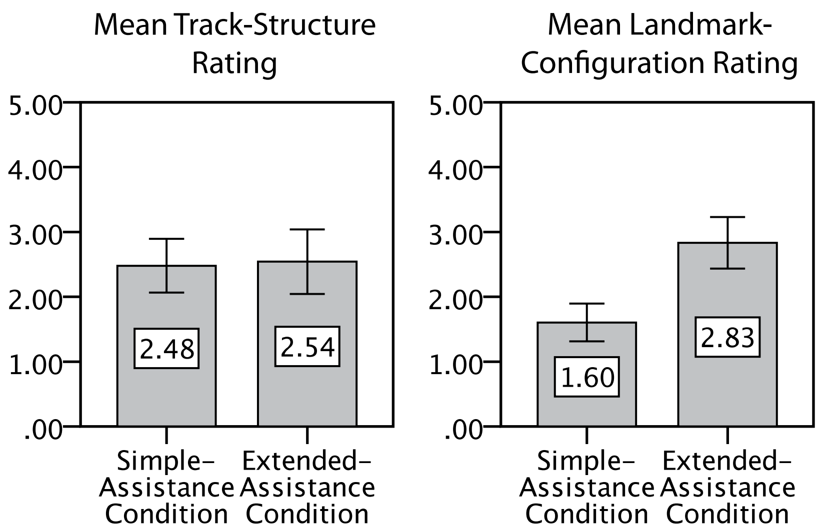


Figure 3.10: Mean ratings of the track structure and the landmark configuration as sketched. Error bars represent the 95 % confidence interval of the mean.

Concerning how well the track structure was reflected in the sketch map (i.e., how well it corresponded to the map learned), no significant effects of assistance condition were present. In line with the results of the configuration questions, the mean ratings of sketch maps produced in the condition in which participants took part firstly ($M = 2.29$, $SE = .23$) are slightly lower than the ones from the condition in which they took part secondly ($M = 2.73$, $SE = .20$). The counterbalanced setting controls for this effect. Like for the analysis of the results of the configuration questions, a paired-sample t-test of equivalences was used to control whether the mean ratings can be considered equivalent. Again, we used a liberal symmetrical equivalence interval of .50 and an alpha level of .05 (two tailed), resulting in the following t value: $t = .24$. This value is lower than the critical constant; therefore, the mean ratings of how well the maps reflect the track structure can be considered statistically equivalent among the conditions.

In contrast, the ratings for how well potential landmarks were reflected in the sketches were significantly affected by the assistance condition. The sketches were better when

[14]I follow the suggestions of Norman (2010) for the evaluation of data based on Likert-type scales.

the map was learned with extended assistance ($F(1,23)=21.39$, $p<.001$, $r=.69$). The difference between the results of the first condition ($M=2.23$, $SE=.21$) and the second condition ($M=2.20$, $SE=.19$) is small.

Recognition Task. As described, aside from the correct part, each part of the recognition task had either an incorrect part of a track or an incorrectly placed landmark, or both. We evaluated separately how well the solution the participants chose reflected the structure of the tracks and the position of potential landmarks. The map was split into quadrants. For each quadrant six options were given: the correct solution and five parts with a wrongly depicted track structure and/or an incorrectly placed landmark. From the number of mistakes, we calculated the number of parts that reflected the correct track structure and the number of parts that reflected correct landmark positions. The mean numbers are shown in Figure 3.11. The figure indicates that participants benefited from extended assistance concerning their acquired knowledge of the landmark configuration. The figure also indicates a slight improvement of participants' knowledge of the track structure.

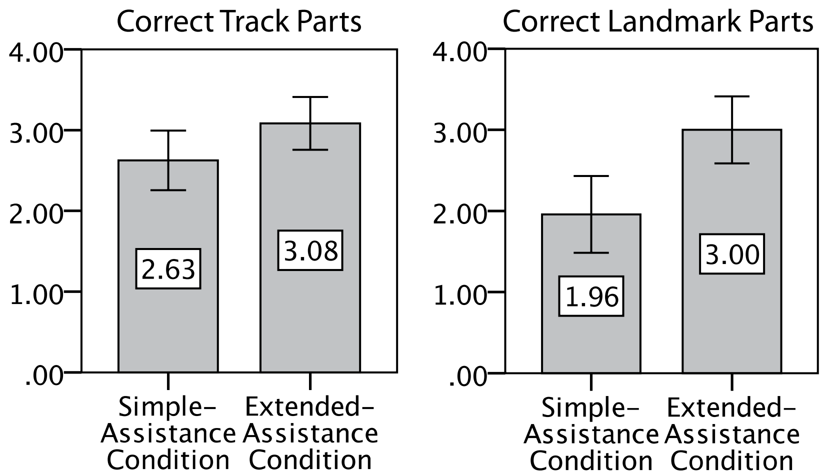


Figure 3.11: Average number of correctly identified parts. Error bars represent the 95 % confidence interval of the mean.

Two separate mixed-design ANOVAs were calculated for the number of parts chosen that correctly reflected the track layout and the number of parts chosen that correctly reflected the landmark layout. Concerning the number of parts that correctly showed the track structure, there were no significant main effects of gender, handedness, which map was used with which condition, the order of the conditions, or which name set was used with which condition. Likewise, there were no significant interaction effects of these variables with assistance condition. Therefore, these variables were stepwise excluded from the model for the further analysis. As in the two other tests, there was no significant effect of assistance condition. However, there was a tendency suggesting that participants performed better under the extended-assistance condition ($F(1,23)=3.87$, $p=.061$). Consequently, the paired-samples t-test of equivalence, using the same equivalence interval (.50) and alpha level (.05) as for the analysis of the other two tasks, resulted in a value that is above the critical constant ($t = 1.97$). This indicates that the number of parts that correctly shows the track structure cannot be considered statistically equivalent. Participants were able to identify the track structure slightly worse in the first condition ($M = 2.83$, $SE = .19$) than in the second condition ($M = 2.87$, $SE = .16$), which is in line with the other tasks.

The number of parts that correctly reflect the landmark layout showed a significant main effect of gender ($F(1, 20) = 8.20$, $p < .05$, $r = .54$). Male participants chose more elements with a correct landmark configuration. Furthermore, there was a significant interaction effect of gender and the order of assistance conditions ($F(1, 20) = 6.39$, $r = .49$, $p < .05$). Males, in contrast to females, performed better when the extended-assistance condition preceded the simple-assistance condition. However, Levene's test indicated that the assumption of equal error variances was broken ($p = .002$). The data showed a significant large effect of assistance condition (taking effects of gender, the ordering of conditions, and their interaction into account) ($F(1, 20) = 7.74$, $p < .05$, $r = .53$). Participants chose more parts with a correct landmark setup after having learned the map in the extended-assistance condition. Participants performed slightly better with respect to landmarks in the first condition ($M = 2.63$, $SE = .35$) they participated in than in the second condition ($M = 2.33$, $SE = .23$) they participated in, which is in line with the other tasks.

3.4.4 Discussion

The experiment was conducted to test whether acquiring knowledge from virtual tactile maps is facilitated by assisting utterances. If assisting utterances facilitate the knowledge-acquisition process, current approaches towards audio-tactile map systems can be improved with respect to their effectiveness.

To evaluate this, participants were asked to explore maps in two experimental conditions: one condition with restricted verbal information (simple assistance) and one condition with additional verbal information (extended assistance). The verbal information included in the latter condition was inspired by the information human assistants gave to map explorers in a corpus study that was conducted previously (see Section 3.2). Three tests were used to assess for spatial knowledge: (1) configuration questions (i.e., participants were asked to answer questions about the layout of objects on the map), (2) sketch mapping, and (3) recognition task (a jigsaw-puzzle-like recognition test).

As expected, the experiment showed that, altogether, participants acquire spatial knowledge more effectively when they receive additional verbal information during tactile map explorations. The results show that knowledge gained from virtual tactile maps consists of two components: knowledge of the structure of tracks (such as streets) and knowledge of the configuration of potential landmarks. We analyzed the data of all tests with respect to the two subtypes. Contrary to our expectations, the overall increase in knowledge was only based on an increase in knowledge of potential landmarks. While the data show a strong increase in knowledge of potential landmarks, participants did not acquire significantly more knowledge of the structure of the track network. With respect to the track structure, a paired-sample t-test for equivalence indicated that the results of configuration questions and sketch maps are statistically equivalent between conditions. This indicates that participants did not integrate the additional verbal information about the track structure in their spatial mental models.

In addition to the main finding that learning virtual tactile maps with additional verbal assistance significantly facilitates acquisition of knowledge about potential landmarks, the study also indicates that all three tests are applicable to assess survey knowledge acquired by tactile map explorations. Configuration questions and sketch mapping have shown a perfect match concerning both statistically significant difference and equivalence. All three tests support the same result, except from the fact that the data for knowledge of tracks of the recognition task are not statistically equivalent between the experimental conditions.

In a similar experiment conducted in the Chinese language, we were able to replicate the most important findings (Lohmann, Yu, Kerzel, Wang, & Habel, 2013). This shows the reliability of the results.

The experiment reported was performed in the context of the development of the VAVETaM system, which is intended to help visually impaired people by automatic generation of verbal assisting utterances for tactile map reading. In Section 3.6, a follow-up experiment with visually impaired people and a similar empirical setting is reported.

Overall, the results encourage the development of the VAVETaM system. However, further research has to investigate the reasons for the absence of an improvement of knowledge of the track structure. The Wizard-of-Oz-like setting can result in a sub-optimal timeliness of the utterances, as it is not possible to make sure that the experimenter is always fast enough with locating and clicking the correct button. Possibly, this suboptimal timeliness of the utterances results in impeded integration of the information given via the verbal channel. However, other possible explanations cannot be ruled out with the current data.

3.5 Validity of Sketch Maps and Configuration Questions

As previously presented (see Section 3.3), the validity of different methods to test human spatial knowledge is frequently discussed in the literature. Research on the methods themselves is regarded essential to get a more complete picture of their strengths and drawbacks (e.g., Kitchin & Blades, 2002).

Experiment 1 was performed to investigate spatial mental models that participants acquired from map explorations with verbal assistance. As suggested in the literature, mutually supportive tests were used, including sketch mapping and answering questions about the configuration of objects to assess the spatial knowledge of participants after they learned the maps. The results of Experiment 1 allow for a comparison of the results of two tasks: configuration questions and sketch mapping (see Section 3.4.2, for a discussion of the tasks). As discussed in Section 3.4.4, on an aggregate level (i.e., on the level of averaged data), the data collected in the two tasks support the same finding.

A comparison of the results is not only possible on an aggregate level, but also on the level of single data points (as suggested by Newcombe, 1985). In the present context, data points are spatial facts (e.g., whether two streets intersect) that are represented in the externalizations of spatial knowledge produced by the participants. Both methods are based on externalizations of spatial knowledge that allow for conclusions about the spatial knowledge of the participant. For example, a sketch map can indicate that according to a participant's spatial knowledge two streets, say A and B, meet in an intersection (which is a spatial fact). In the same manner, the answer to a question like 'Do streets A and B form an intersection?' allows for conclusions about the spatial knowledge of the participant. If the two methods agree in the conclusions about spatial knowledge they allow for, this is further evidence for the validity of the methods: The two methods test the same thing, which then plausibly is spatial knowledge.

3.5.1 Method

The data were collected in Experiment 1. Detailed information about participants, material, and testing methods is provided in Section 3.4.2. The data of the configuration-questions task and of the sketch-mapping task offer the possibility for a direct comparison.

In the experiment, participants learned virtual tactile maps. To facilitate learning, they received different styles of assisting utterances. After each learning condition, participants answered questions concerning the spatial layout and were asked to produce a sketch map.

For the purpose of comparison, the results of these two tests were evaluated in the following way: Each of the sketches was analyzed by the researcher. If the sketch clearly reflected a spatial fact that was also asked for in one of the configuration questions, the answer to that question as indicated in the sketch was noted in a table. After this was done for all sketches, we counted whether the answer to the questions based on the spatial layout of the sketch agreed with the one the participant gave to the configuration questions.

The following example illustrates the evaluation method. Question 19 (see Table 3.11) addresses whether a street called 'Amselweg' intersects with a street called 'Blumenstraße'. In the sketch map shown in Figure 3.12, these two streets do not intersect. In the figure, the relation is indicated by the arrow labeled '19'. As shown, the participant did not draw the streets intersecting. A visualization of the original map is shown in Figure 3.13. As can be seen, in the original map the streets do not intersect. If the sketch map and the questions both reflect the spatial mental model in an equal fashion, it is reasonable that the same participant also answered 'no' to question 19. If the participant did answer with 'no', we counted this pair of data points as equal. Pairs of data points with questions that were answered with 'do not know' were omitted from the evaluation.

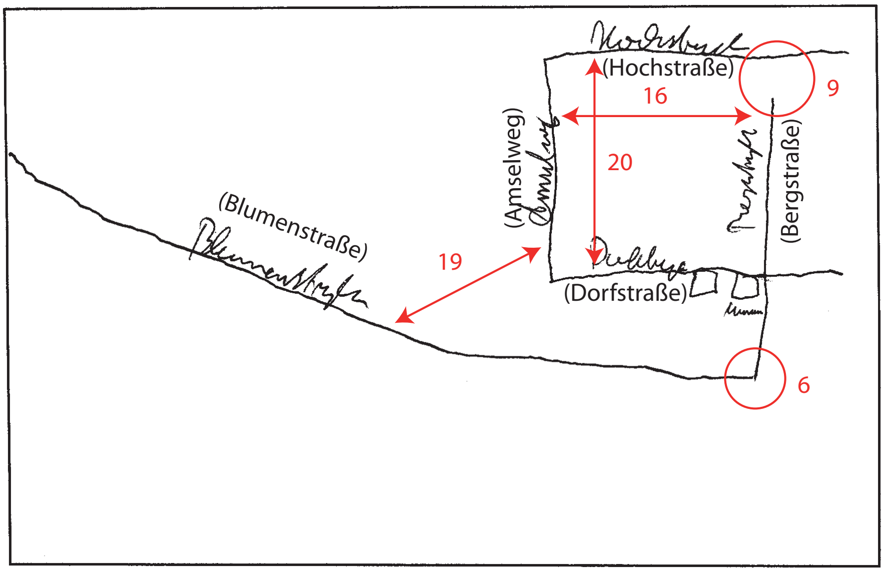


Figure 3.12: Scan of a sketch map produced by a participant, augmented by additional information. Data points that reflect knowledge that was also asked for in the configuration questions are indicated by arrows and circles with the number corresponding to the question (see Table 3.11). A visualization of the original tactile map is shown in Figure 3.13.

Table 3.11: Translations of some of the questions posed in the experiment. The numbers correspond to knowledge reflected at the highlighted and numbered parts in Figure 3.12.

Question Number	**Translation of the Question**
6	Do Blumenstraße and Bergstraße form a T-intersection?
9	Do Hochstraße and Bergstraße intersect?
16	Do Bergstraße and Amselweg form a T-intersection?
19	Do Amselweg and Blumenstraße intersect?
20	Do Hochstraße and Dorfstraße meet?

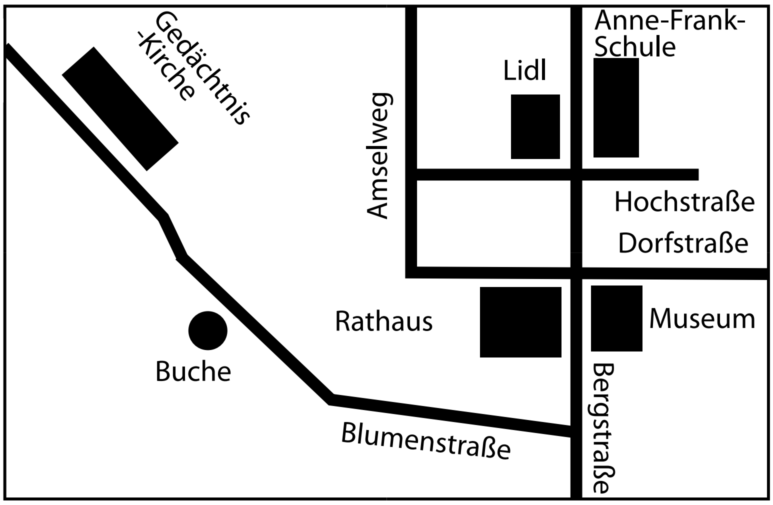


Figure 3.13: The original map for the sketch shown in Figure 3.12.

3.5.2 Results

Applying the procedure described above resulted in 504 pairs of data points that were compared. The data points agree to 87.70 %, which is a fairly high agreement ($\Phi = .51$, $\Phi_{max} = .97$, $\Phi/\Phi_{max} = .53$[15]).

3.5.3 Discussion

The results of two tests of spatial knowledge used in Experiment 1 were compared. This comparison was performed on the basis of individual data points.

The results indicate that both methods are valid measures of humans' spatial knowledge. An agreement of 87.70 % of the paired data points can be regarded as high agreement, taking into account that participants might have guessed about the spatial layout in both tasks when they were not sure about how to answer the configuration questions or how to sketch a part. This result supports the interpretation that both, sketch maps and configuration questions, are valid assessment methods to be used in measuring the accuracy of spatial mental models of sighted people—for example, to compare the performance of different maps or geographic information systems.

It can only be speculated about the source of the remaining variation between the methods. Even though guessing is a likely cause, other explanations that take the demands of the tasks into account are possible. The reported result has to be seen in the context of the fact that in the configuration-questions task coarse-grained, qualitative knowledge was needed to answer the questions posed and the maps used in the experiment were relatively sparse. Therefore, the comparison reported here is an evaluation

[15]Note that the Φ/Φ_{max} value is not robust in the case of the data presented due to extreme marginal values in the contingency table (Davenport & El-Sanhurry, 1991).

of the degree to which the two methods can be used to measure coarse-grained spatial survey knowledge. Used in this context, the effects of individual sketching ability and the different demands of the tasks do not affect the validity of the methods to a high degree.

3.6 Experiment 2: Visually Impaired Participants

3.6.1 Introduction

As discussed in the introduction of this chapter, Experiment 2 was designed as a follow-up experiment to Experiment 1. The basic idea was to repeat Experiment 1 with visually impaired people, who are the potential user group of VAVETaMs. The experiment was conducted to investigate whether VAVETaMs are suitable for visually impaired users. Furthermore, the experiment shows whether the results of Experiment 1 can be conferred to the group of potential users.

The data of seven visually impaired participants were evaluated. As in Experiment 1, we asked participants to explore virtual tactile maps in two assistance conditions: simple assistance and extended assistance. The simple assistance condition was the baseline condition for the study. In this condition, only named objects were identified verbally. In the extended-assistance condition, more information was included in the verbal assistance (e.g., information about relations between objects and identification information for objects without a name).

In Experiment 1, three tasks were used to test participants' spatial knowledge: configuration questions, sketch mapping, and recognition task. Obviously, ordinary sketch mapping and a visual recognition task are not suitable testing procedures for visually impaired people. Therefore, we developed similar tasks that are based on touch. As the usability of tactile sketch mapping and tactile recognition tasks to test visually impaired people is not well-covered in the literature, the use of this tasks was exploratory (see Section 3.3, for a literature review).

3.6.2 Method

Participants: Blind and Visually Impaired Participants

Initially, nine visually impaired participants took part in the experiment. The data of two participants was excluded from further evaluation, due to problems that occurred in the experimental procedure (one participant broke up the exploration of the map too early and one participant's phone rang extensively in one of the conditions). This results in the evaluated data of seven participants (in research with visually impaired participants, comparable small numbers of participants are common, see Lawrence, 2011).

When selecting participants, we intentionally did not restrict for either congenitally blind, late blind, or visually impaired people with some remaining sight, as VAVETaMs are designed to be an appropriate means for individuals of all of these groups.

All participants were compensated on a monetary basis. A part of the experiments (three participants) took place at two different quiet rooms at the Louis Braille center in Hamburg. Five (the data of four of those were evaluated) participants traveled independently to our laboratory at the University of Hamburg. One experiment was performed by a participant at home (excluded from data analysis because the phone rang extensively in one of the conditions). A participant with remaining sight for contours was blindfolded to ensure that he or she did not use the visual sense. All participants were naïve about the purpose of the experiment and had not used the device, explored the maps, or performed the experimental tasks prior to the experiment. They gave informed consent, reported to speak German on a native speaker level, and reported to have no cognitive disorders. One participant was a non-native speaker but lived for 30 years in Germany and spoke German on a native-speaker level. Participants had a mean age of 41.86 years (*SD* 9.76 years). Five participants were male, two female. The participants' age at the time of participation, the cause for their visual impairments, and its the time of its onset are given in Table 3.12. The table is based on participants' self reports.

Table 3.12: Overview of the participants of Experiment 2 (data of participants 7 and 9 was excluded from evaluation).

Participant	**Age**	**Onset Age of Impairment**	**Cause for Visual Impairment**	**Remaining Sight**
1	49 years	Congenitally	Visual nerve not developed	Dark/Light
2	57 years	51 years	Glaucoma	Contours
3	37 years	30 years	Innate impairment	None
4	42 years	6 years	Brain tumor	None
5	34 years	0.5 years	Eye cancer	None
6	48 years	Congenitally	Retinopathia praematurorum	Dark/Light
7	-	-	-	-
8	26 years	Congenitally	Leber's congenital amaurosis	None
9	-	-	-	-

Materials and Procedure

Material. The maps that participants explored were the same as used in Experiment 1. Two virtual tactile maps with an equal number of tracks (5 tracks) and potential landmarks (6 potential landmarks) were used (see Section 3.4.2, for a discussion).

Like the maps, the pre-recorded assistances were the same as used in Experiment 1. The experiment was performed as a Wizard-of-Oz-like experiment. The playback of the assisting utterances was controlled by the experimenter using the custom-built interface discussed in Section 3.4.2.

The participants explored the maps, in accordance with Experiment 1, receiving different styles of assisting utterances. In the simple-assistance condition, they only received information about the names of entities. In the extended-assistance condition, participants received additional verbal information. For example, they were informed about spatial relations between map objects. Refer to Section 3.4.2 for an exhaustive description.

Testing Methods. In Experiment 1, three mutually supportive tests for spatial knowledge were used: (1) configuration questions, (2) sketch mapping, and (3) a recognition task. One of the goals of Experiment 1 was to investigate whether all test methods support the same conclusions, which they did. This indicates that each of the tests can be used validly in similar experimental setups with sighted participants.

The configuration questions were suitable to test visually impaired people. This method was used without adaption, except from a minor modification of question 5: In Experiment 1, this question was excluded from further evaluation as it was wrongly formulated. The question did not exactly match in the different sets of questions for the two maps and the two name sets (see Appendix B.4 for the sets of questions used). In the currently discussed experiment, question 5 was modified to match exactly among the sets.

Obviously, visual sketch mapping and solving the visual recognition task were not applicable to test the spatial knowledge of visually impaired people. Therefore, these methods were adapted to touch-based versions. Sketch mapping was done using a tactile drawing pad called 'Draftsman Standard' (see Figure 3.14). The tactile drawing pad works as follows: A special foil is fixed on a rubber surface. This foil raises when lines are drawn with a common ball pen. The raised lines can be felt via touch. Thus, they provide orientation with respect to what has already been drawn for the visually impaired user. Additionally, the pen creates a visual line as it does on usual paper, which supports analysis. In Experiment 1, participants wrote written labels with the names of the entities next to the map objects they drew. To adapt for that, visually impaired participants were told that they could point on map objects and call out the objects name. The experimenter then added a written label in alphabetic writing to the map object. Most visually impaired participants either reported to know this kind of sketching tool from school or were quickly able to understand it. Yet, a commonly discussed problem of sketch-mapping techniques, namely that they rely on sketching experience (see, e.g., Newcombe, 1985; see also Lohmann, 2011, and Section 3.3 for overviews), is intensified because sketching is not a well-trained every-day task for visually impaired people. According to Golledge (1976), "[a]sking an individual with limited graphic representational capabilities to construct a map of the environment severely limits his or her ability to show the extent of knowledge [...]" (p. 308). On the other hand, Jacobson (1992, 1996) reports on a successful use of sketch mapping on a drawing board with visually impaired people.

Blind sketching is a task that is more demanding than sighted sketching. Consequently, the sketch maps that sighted participants produced and those that visually impaired participants produced are not comparable. Besides that, it is questionable whether the tactile sketching task is sensitive enough to detect differences in spatial knowledge caused by the condition. However, in a repeated-measures design, the problem of individual graphic abilities is compensated. Therefore, as long as graphic ability suffices to produce sketches that can be used as an indicator of spatial knowledge at all, the results can be considered valid.

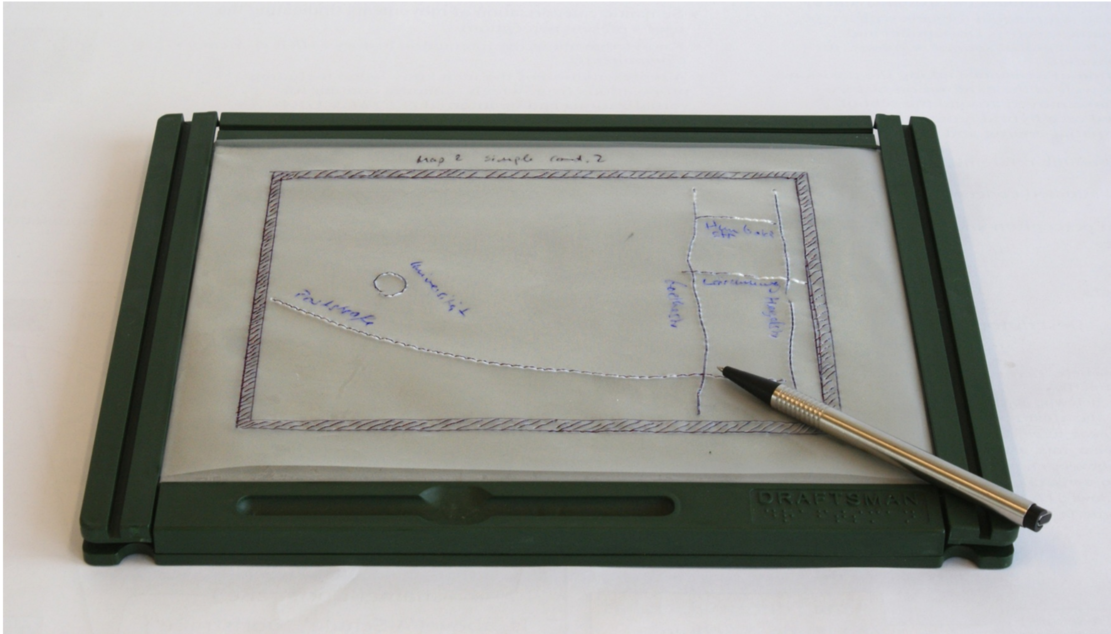


Figure 3.14: The Draftsman Standard sketching pad for visually impaired people.

Similar arguments affect the adaption of the visual recognition task. Like sketch mapping, tasks that are similar to jigsaw puzzles are not everyday tasks for visually impaired people. While most sighted people have experience with jigsaw puzzles (at least from solving them in their youth), most visually impaired people (at least with an early onset) do not. In order to adapt this test, we used medium-density fiberboard plates and a model-making router to cut out indentations as in the virtual tactile maps. The layout of the parts was identical to the layout of the visual parts. In the visual version of the recognition task, participants were asked to choose among six options for each of the quadrants, resulting in 24 options altogether. Preliminary tests showed that this set of options was too large for a non-visual task, even though the goal position and the orientation of the parts were given. Consequently, only options for one quadrant of the maps were given. For each of the two maps the options to choose upon were given for a different quadrant. This was done to avoid that participants intensively learned only one quadrant of the map in order to improve the results in the recognition task. For Map 1, the part for which the options were given was the upper right part. For Map 2, the options were given for the bottom right part. Participants were not informed for which part of the map the options would be given before they solved the task. Figure 3.15 shows a picture of the tactile recognition task for the first map.

While the reduction of choice options to one quadrant was necessary to make the task feasible, it entailed the problem that the number of possible errors was largely reduced from eight to two: In the visual version, choice options with two possible mistakes were given for all four quadrants and in the tactile version only for one quadrant. Therefore, unsystematic variance due to guessing increased largely in comparison to the range of possible results.

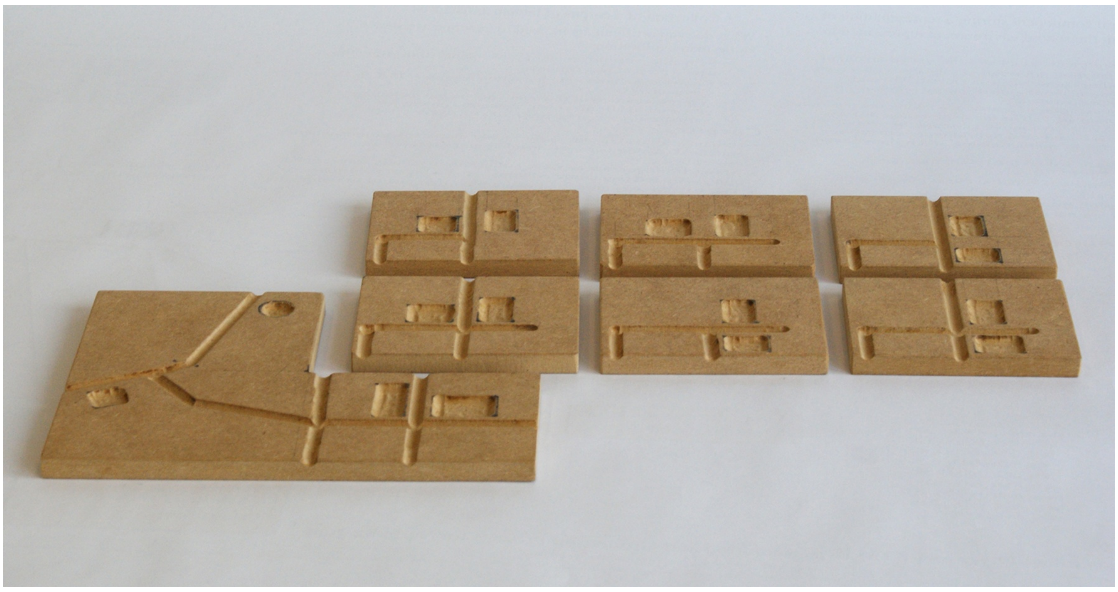


Figure 3.15: The tactile version of the recognition task. In the tactile recognition task, choice options were only given for one quadrant.

The use of tactile sketch mapping and the tactile recognition task was exploratory. The literature did not allow to anticipate whether these tasks are feasible. Yet, including them in the experimental procedure was necessary. We wanted to repeat Experiment 1 as completely as possible to exclude unwanted biases on the result. Usually, spatial knowledge acquisition is a task-driven activity. Humans acquire spatial knowledge for a reason (cf. Avraamides et al., 2004, for an example that instructing participants can affect spatial mental models acquired). As reported, both experiments were performed as repeated-measures designs and participants were informed about the tasks they had to solve after the learning phase. It is possible that participants focus on specific aspects of spatial knowledge when they know which tasks they have to solve. Consequently, non-visual adaptions of all assessment tasks that were used in Experiment 1 made sure that no change in the experimental procedure influenced the results.

Procedure. As described, basically, the procedure that was applied in Experiment 1 was repeated. The experiment was performed as repeated-measures experiment with the assistance condition as within-subject variable. This design was chosen to overcome problems of individual differences in the ability to understand virtual tactile maps resulting from differences in spatial abilities biasing the results.

To avoid carry-over effects, two different maps and name sets were used for the two experimental conditions (as in Experiment 1). The second condition was always performed with the map, the assistance condition, and the name set that were not used in the first condition. The experiment was randomized concerning the order of assistance conditions, maps and condition combination, and name set and condition combination.

To ensure that participants understood interaction with the experimental system, they were trained in different aspects of it. Firstly, they were introduced to the haptic interface by explanations of the experimenter and explored the device itself haptically. Then, they were interactively introduced in virtual tactile maps with verbal assistance. In this training, they explored a map, in the first phase assisted by the experimenter who gave assisting utterances and in the second phase by pre-recorded assistances controlled by the experimenter. Sighted participants in Experiment 1 were allowed to see a visualization of the virtual tactile map used in the first training phase after some time of haptic interaction. To create comparable conditions in Experiment 2, a physical copy of the training map was created. Participants were allowed to explore this physical copy in the first training phase. The physical copy was manufactured in the same way like the tactile recognition task.

The map used for training purposes was different from the two maps used in the experimental conditions. In the training procedure, the map objects had artificial names such as ‘A-building’ or ‘alpha-street’. To conclude the training, participants were tested for their ability to identify the shape of objects (such as a triangle and a square) presented as indentations (like the potential landmarks) in virtual tactile maps. Additionally, they were asked to follow a complex track structure without leaving it while receiving and following assistance given by the experimenter. The training procedure took 35–60 minutes, including the training test.

After a short break, the experimenter read the instructions out aloud. The instructions included information about the time participants had for exploration and about the tasks they had to solve after learning. The tasks were identical in both conditions; accordingly, participants would have known which tasks they had to solve anyway after they had performed the first condition. Hence, information about the tasks was included to minimize order effects resulting from knowledge of the tasks. Participants were instructed to read the map in a way that they would find the route from each potential landmark to each other. Each map exploration was limited to eight minutes. After the map explorations, first the configuration-questions task was performed, then participants created the tactile sketch map, and finally did the tactile recognition task. A break (about ten minutes) was made before participants performed the second exploration with the same time constraint as the first exploration. In general, instruction, material, and procedure were like in Experiment 1, aside from the tactile adaptions of the tasks and of some of the material used in training, which were described. The experiment took about two hours including the training session. After the procedure as described was performed, an additional interview and system evaluation on a Likert-type scale was conducted.

3.6.3 Results

Configuration Questions. Overall, visually impaired participants were able to answer more questions correctly when they learned the map in the extended-assistance condition ($M = 11.86$, $SE = .96$) than when they learned it in the simple-assistance condition ($M = 7.43$, $SE = 1.15$). See Figure 3.16 for a visualization. A paired t-test was calculated showing that the result is significant and the effect is large ($t(6) = 5.10$, $p = .002$, $r = .90$). The differences between the overall number of correct answers in the first condition ($M = 9.85$, $SE = 1.44$) and in the second condition ($M = 9.43$, $SE = 1.34$) are small.

Experiment 1 showed that blindfolded sighted individuals improve only in their knowledge acquisition about potential landmarks, not about tracks. This finding was supported by a principal component analysis. Whereas the questions involving potential landmarks (questions 1–4[16] and 10–15) loaded highly on one component, the other questions (with the exception of question 8) do load highly on a second component. Therefore, the first component was called 'landmark knowledge' and the second component was called 'track knowledge'.

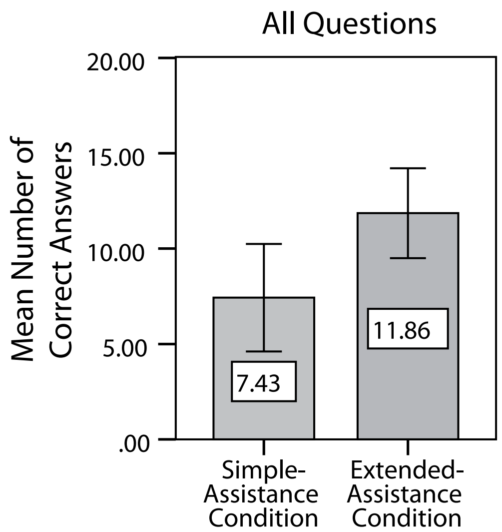


Figure 3.16: Mean number of correct answers to all questions of the configuration-question task. Error bars represent the 95 % confidence interval of the mean.

With the data of the experiment presently discussed, the finding that the configuration questions form a two-dimensional scale could not be replicated in a principal components analysis. However, the data were analyzed in the same dimensions as in Experiment 1 to enable comparison between the results. The procedure described in Section 3.4.3 was applied. The results were analyzed separately for track-knowledge questions and

[16]Question 5 was excluded in Experiment 1.

landmark-knowledge questions. Figure 3.17 shows the mean number of correctly answered questions to each of the set of questions in each of the conditions.

Separate paired t-tests were calculated for the track and the potential-landmark questions. Due to the small sample size, controlling for handedness and gender effects and effects of the order of condition, the maps used with which condition, and the name set used with which condition was not possible.

The number of correctly answered track questions increased significantly from simple-assistance condition ($M = 6.00$, $SE = .79$) to extended-assistance condition ($M = 8.14$, $SE = .46$, $t(6) = 2.91$, $p = .027$, $r = .76$) and the effect is large.[17] Participants were able to answer significantly more questions correctly when they learned the map with extended assistance. The mean values of the first condition in which participants took part ($M = 7.00$, $SE = .69$) and the second condition ($M = 7.14$, $SE = .85$) in which they took part differ only slightly.

Even though the number of correctly answered questions involving potential landmarks increased when the map was learned with the extended-assistance condition ($M = 3.71$, $SE = .97$, $t(6) = 2.15$, $p = .075$, $r = .66$) compared to learning with the simple-assistance condition ($M = 1.43$, $SE = .75$), it did not increase significantly. There is a small difference between the results of the condition in which participants took part firstly ($M = 2.86$, $SE = .85$) and the condition in which they took part secondly ($M = 2.29$, $SE = 1.08$).

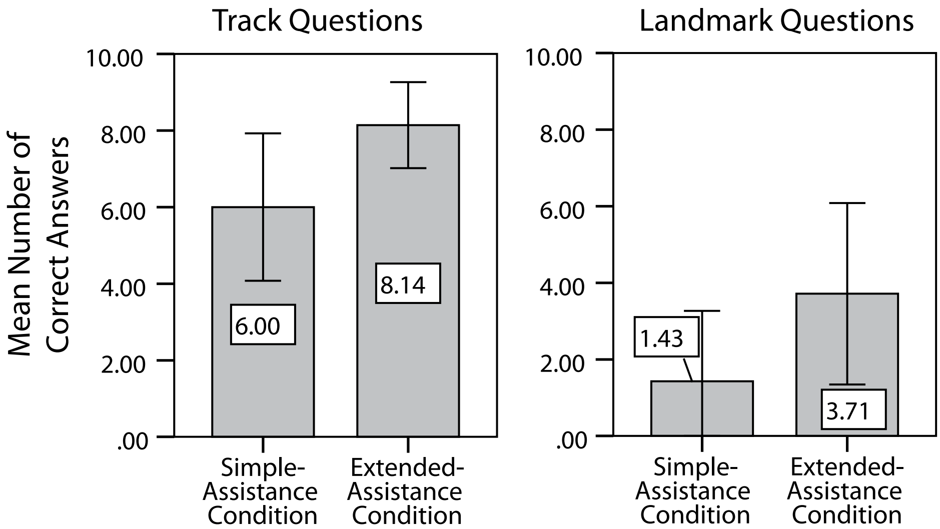


Figure 3.17: Mean number of correct answers to the track questions and the potential-landmark questions of the configuration-question task. Error bars represent the 95 % confidence interval of the mean.

[17] All values reported are based on two-sided testing.

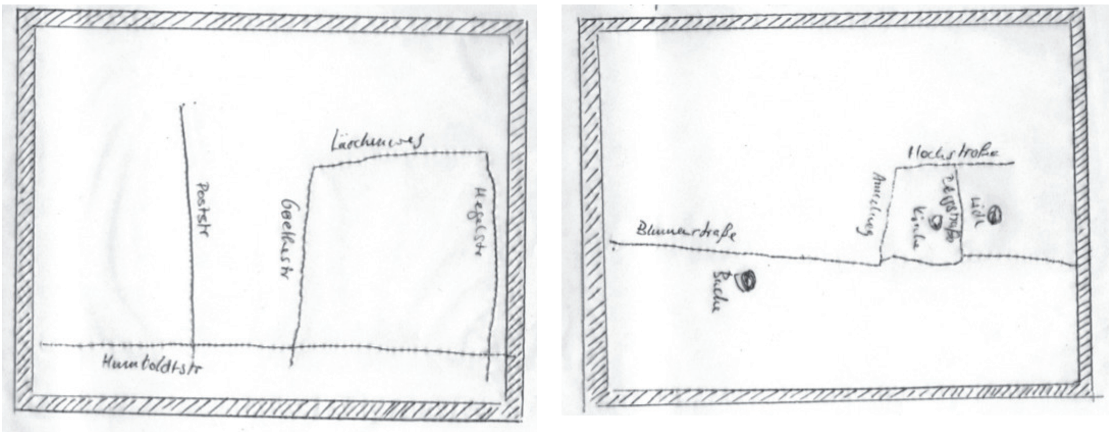
Tactile Sketch Mapping. As the sketch maps collected in Experiment 1, the sketch maps collected in the present experiment were rated by two independent raters naïve to the condition under which they had been produced. The sketch maps were rated in two dimensions, namely how well the structure of tracks is reflected in comparison to the original map and how well the configuration of potential landmarks is reflected in comparison to the original map. The raters gave a score on a 1–5 Likert-type scale for each or the sketches. To test the raters for agreement, the intraclass correlation coefficient (ICC) was calculated separately for the track-structure ratings and the potential-landmark-configuration ratings. Both show a reasonable agreement (track-structure ratings $ICC(3,1) = .80$, potential-landmark-configuration ratings $ICC(3,1) = .91$). Ratings were averaged for further analysis.

Statistical analysis did not show any significant effects of assistance condition in both, the ratings of the track structure and the ratings of the landmark configuration. The mean rating of the track structure on sketches that participants produced after they received simple assistance is slightly lower ($M = 2.29$, $SE = .44$) than on the sketches produced after the extended-assistance condition ($M = 2.50$, $SE = .41$, $t(6) = .89$, $p = .41$, $r = .34$). The same holds true for the ratings of the landmark configuration (simple assistance: $M = 1.64$, $SE = .24$; extended assistance: $M = 2.07$, $SE = .38$, $t(6) = 1.22$, $p = .27$, $r = .44$). As the mean rating of the landmark configuration indicates, participants sketched slightly better in the second condition they took part in (first condition: $M = 1.79$, $SE = .21$; second condition: $M = 1.93$, $SE = .41$). The same pattern can be found in the mean ratings of the track structure (first condition: $M = 2.21$, $SE = .39$; second condition: $M = 2.57$, $SE = .46$).

Generally, a high variance of the quality of sketches among participants was observed. Figures 3.18 and 3.19 show examples for the raised-line sketch-map drawings participants produced in the experiment. The figures show sketches that two of the visually impaired participants produced after the two experimental conditions. The two sketches in each figure were produce by the same participant.

Validity of Tactile Sketch Maps. In Section 3.5, I discussed an evaluation whether sketch maps and configuration questions can be considered valid methods to assess humans' spatial knowledge. Basically, the agreement between configuration questions and sketch maps was examined. We repeated the evaluation as described with the data from the current experiment. The configuration questions and the sketch maps agree to 70.97 percent ($\Phi = .45$, $\Phi_{max} = .72$, $\Phi/\Phi_{max} = .63$[18]). In the dataset that was collected in Experiment 1 with sighted participants, the two methods agreed in 87.70 percent of all cases. The lower agreement between the testing methods in Experiment 2 indicates that sketching ability has a larger effect on the results when visually impaired people are

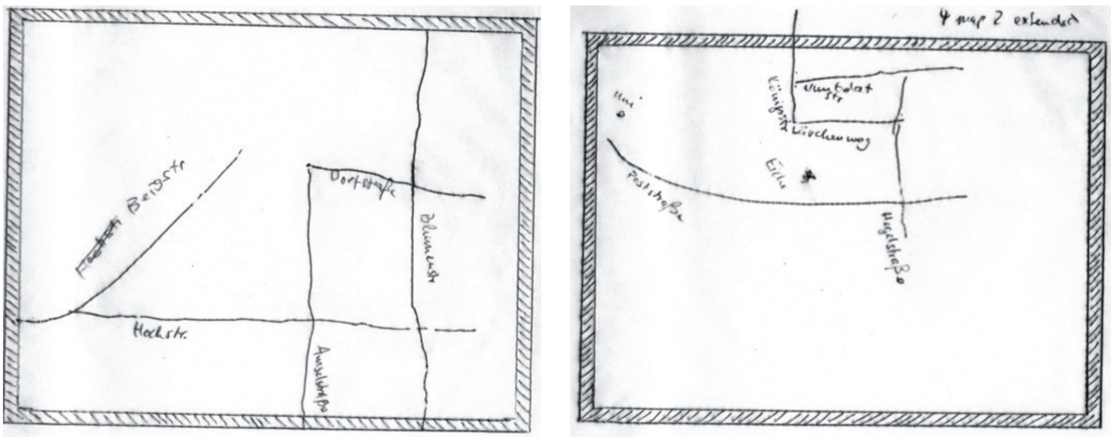
[18] This value is higher than the value presented in Section 3.5. Yet, as discussed there, the value is not robust against the case of extreme marginal values of the contingency table, which leads to biased results (Davenport & El-Sanhurry, 1991). Such extreme marginal values are present in the data discussed.



Original Map: Map 1, simple condition

Original Map: Map 2, extended condition

Figure 3.18: Example of a relatively low-rated pair of sketches of one of the visually impaired participants.



Original Map: Map 1, simple condition

Original Map: Map 2, extended condition

Figure 3.19: Example of a relatively high-rated pair of sketches of one of the visually impaired participants.

tested—at least assuming that configuration questions are a valid method, which is not affected by visual status.

Tactile Recognition Puzzle. The recognition-task data was calculated in the following way: The actual number of mistakes was subtracted from the number of mistakes possible. For the overall score, this calculation leads to two possible mistakes (one mistake with respect to the track structure and one with respect to the landmark configuration). The tactile recognition task does not indicate any significant effect of assistance condition. Overall, the scores are slightly lower when participants received extended assistance ($M = 1.29$, $SE = .18$) than when they received simple assistance ($M = 1.43$, $SE = .20$, $t(6) = .55$, $p = .60$, $r = .22$). The scores for how well the track-structure is reflected are slightly higher when participants receive extended assistance ($M = .85$, $SE = .14$) than when they receive simple assistance ($M = .57$, $SE = .20$, $t(6) = 1.00$, $p = .36$, $r = .38$). Contrary, participants chose less correct parts concerning the landmarks in the extended-assistance condition ($M = .43$, $SE = .20$) than they chose in the simple-assistance condition ($M = .85$, $SE = .14$, $t(6) = -2.12$, $p = .078$, $r = .65$). Altogether, the data is not interpretable due to the small range of possible outcomes, the large amount of unsystematic variance compared to systematic variance and the small number of participants. In the experiment, a number of participants indicated that they were guessing in this task.

3.6.4 Discussion

The experiment reported was a follow-up experiment to Experiment 1, in which we tested whether blindfolded sighted people benefit from extended verbal assisting utterances when they learn a virtual tactile map. Again, the effect of two different styles of assisting utterances was compared. The procedure was like in Experiment 1. Two of the tests of spatial knowledge were adapted for testing visually impaired people. The inclusion of these tasks was important to ensure that the goal of map learning did not change in comparison to Experiment 1. Altogether, sketch mapping was applicable, but there is a high variance between the participants. Furthermore, the effect of assistance condition is only small, which questions the sensitivity of sketch mapping in experiments with visually impaired participants.

Even though participants reported that they considered the tactile recognition task understandable (see Section 3.6.5), they frequently indicated that they were guessing while solving the task in the experiment. Choice options could only be given for one quadrant so as to make the task applicable. This resulted in a maximum of two correct choices. As each incorrect part resulted in one or two mistakes, guessing resulted in a comparably large unsystematic variance. Hence, the sensitivity of the tactile version of the recognition task to detect differences in spatial knowledge, especially given the small number of participants, is restricted. Consequently, the data from the tactile recognition task cannot be interpreted.

The tactile sketch-mapping task does not show significant effects of assistance condition. Still, it supports the major finding of the configuration-questions task: Overall, extended assistance leads to a significant increase of effectiveness of the knowledge-acquisition process in comparison to simple assistance. Furthermore, in contrast to Experiment 1, participants gained more knowledge about tracks when they learned the map with extended assistance than when they learned the map with simple assistance.

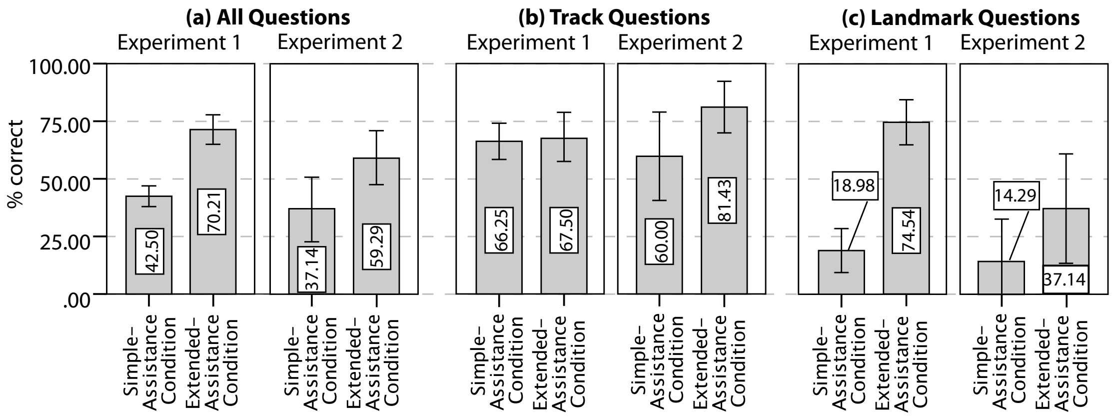


Figure 3.20: Comparison of the results of the configuration-questions task between Experiments 1 and 2. Results are presented as percentage of correct questions. Error bars represent the 95 % confidence interval of the mean.

The configuration-questions task and the exploration and training were exactly like in Experiment 1 (except from a minor modification of question 5, as discussed in Section 3.6.2). Therefore, it is possible to compare the results of visually impaired and blindfolded sighted participants in this task. Altogether, visually impaired participants acquired slightly less knowledge of the map (Figure 3.20 (a)), which is possibly due to more expertise of sighted participants with map-like representations. Visually impaired participants acquired more knowledge about the track structure than sighted participants in the extended-assistance condition, but less in the simple-assistance condition (Figure 3.20 (b)). In both conditions, visually impaired participants acquired less knowledge about potential landmarks (Figure 3.20 (c)).

As the configuration-questions task shows, overall, visually impaired people benefit significantly and to a large degree from receiving extended assistance in comparison to simple assistance. To rely the conclusions on the configuration-questions task is justified for several reasons. This task was shown to be a valid measure for the quality of the spatial mental model of participants in Experiment 1 (see Section 3.4 and Section 3.5). Sketch-mapping data provides non-significant support for this finding. The adapted tasks—that is, sketch mapping and recognition task—are novel tasks for most visually impaired participants. In comparison, the visual versions used in Experiment 1 are well-learned familiar tasks for sighted people. Due to the large unsystematic variance (i.e., variance that cannot be explained by a manipulation of independent variables) of the tactile recognition task, the data of this task cannot be interpreted.

Visually impaired people benefit from extended assistance with respect to their knowledge of the track structure and of the landmark configuration. Visually impaired people are the potential user group of VAVETaMs. Therefore, the results indicate the appropriateness of the information included in the set of assisting utterances. This result shows the capability of natural language to communicate spatial knowledge non-visually in a system combining virtual haptics and natural language. Thus, the results of the experiment encourage to develop the VAVETaM system.

3.6.5 Interview and System Evaluation

Interview: System Evaluation and Identification of Exploration Strategies

Participants were interviewed about the interaction with VAVETaMs after they took part in the experiment. The interviews were conducted in a semi-structured way basically covering two topics: the strategies participants used for acquiring knowledge from the map and whether they considered the system helpful (see Appendix B.7, for the interview guideline). Especially, the interview focused on the participants' attitude towards the assisting utterances that were given in the extended-assistance condition. One participant could not be interviewed due to time constraints (the experiment took longer than average).

The first participant of the experiment was a blind professional (H.) for evaluating assistance technologies. Therefore, the interview with H. was structured more open than the interviews with the other participants and more centered on the evaluation of the interface than on individual strategies. H. stated that some blind people use maps extensively; for example, a visually impaired friend of H. uses sea charts and a speaking compass for sailing. Furthermore, H. reported that blind orientation is mainly focused on audio information. For example, trees, hedges, streets, audio enabled traffic lights, and buildings emit sound or reflect the sound of the white cane in a way that helps for orientation. In orientation and mobility training, visually impaired individuals are trained to acoustically detect and count entrances to buildings to use this information for navigation. Overall, H. considered the virtual haptic maps and language assistance understandable and very helpful if used to acquire an overview of a spatial environment. H. suggested that when the system is built, possibilities to increase the level of detail should be explored.

All other participants regarded the assisting utterances they received in the extended-assistance condition as helpful, as well. They considered the utterances suitable with respect to the information included in the verbal assistance and the way information was expressed.

Exploration strategies are a likely candidate for individual differences in the acquisition of spatial knowledge from tactile maps. Past research with visual maps (e.g., Thorndyke & Stasz, 1980; MacEachren, 1992) and traditional physical tactile maps (Bérla et al., 1976; Ungar et al., 1995, 1997; Blades et al., 1999; Perkins & Gardiner, 2003) has shown that different knowledge acquisition strategies can account for interpersonal differences in the effectiveness of knowledge acquisition.

Four of the five participants who were interviewed specifically for their exploration strategy gave answers that allow for a classification of the strategy they used for explo-

Table 3.13: Translations of some phrases that were used to identify the fix-object strategy and the overview strategy. The original German phrases are given in brackets.

Fix-object strategy	**Overview strategy**
• Salient street structures [‘Markante Straßenmuster’] • Orientation points [‘Orientierungspunkte’] • Known points [‘Bekannte Punkte’] • Homewards [‘Nach Hause’] • Go back [‘Zurückzugehen’] • Not to loose the thread [‘Nicht den Faden zu verlieren’] • In a piecemeal fashion [‘Stück für Stück’] • First, then back [‘Erstmal, dann zurück’] • . . . , to really memoryize them [‘. . . , diese wirklich einzuprägen’] • Inspect unclear parts [‘Unklare Stellen untersuchen’] • Immediate surrounings [‘Unmittelbare Umgebung’]	• To get an overview [‘Überblick zu verschaffen’] • First, find all streets [‘Erst einmal alle Wege finden’] • Overall [‘Im Groben und Ganzen’] • How many streets there are [‘Wie viele Straßen es gibt’] • Straight forward to the next. . . [‘Direkt weiter in die nächste. . . ’] • Explore what is on the map [‘Erkunden, was es gibt’] • All streets [‘Alle Wege’] • To explore everything [‘Alles abzulaufen’] • From the very X (left, top, . . .) to the very Y (right, bottom, . . .) [‘Einmal von ganz X (links, oben, . . .) nach ganz Y (rechts, unten, . . .)’] (this phrase does not indicate the overview strategy when exploration of one map object is described)

ration. The description of the strategies allows for a classification of how participants organized their knowledge acquisition process (see later in this section for a detailed discussion of this analysis). The strategies described can be grouped to two different classes.

Three participants reported that they used a strategy that can be called *'overview strategy'*. This strategy is characterized by the approach to firstly acquire an overview of which objects are on the map and then focus on their positions, shape, and their relations. Two other participants reported that they used a strategy that can be called *'fix-object strategy'*. This strategy is characterized by organizing new map objects piece by piece in relation to a salient known one.

The data do not show any evidence that the choice of exploration strategy affects the effectiveness of the knowledge acquisition process. However, due to the small number of participants, this lack of an effect should not be interpreted as evidence against an effect of strategy on the knowledge acquisition process. Nonetheless, a similar finding is reported by Lahav and Mioduser (2008b) who were, as well, able to identify different strategies in audio-tactile virtual-environment explorations (in their case, the virtual environment is based on a force-feedback joystick and used to present indoor environments; hence, the categories cannot be compared). Similar to our result, the authors did not find significant differences in performance among the groups that used different exploration strategies.

To evaluate the interviews, all participants interviews were transcribed. We analyzed the transcriptions for information about map-exploration strategies. For this purpose, a set of key phrases that are indicators for one of the strategies was created. Table 3.13 shows these key phrases and the associated strategy. The collection of key phrases is based on the data of Experiment 2 and Experiment 3 (as reported in Section 3.7, in Experiment 3, participants were interviewed as well). We categorized the key phrases in the following manner: We first listed all the phrases that we considered being concerned with exploration strategies and then grouped them according to the strategies they describe.[19] If a participant used only one of the key phrases in the interviews, this was not counted as evidence for the use of one or of the other map-exploration strategy. The interviews were manually classified and only when the strategy was stated clearly (that is, by consistent and repeated use of phrases similar to the key phrases shown in the table), we supposed that the participant used one of the strategies. See Appendix B.8 for a translation of the transcript of an example interview.

System and Task Evaluation

Subsequently to the interview, participants were asked to indicate agreement to 7 statements on a 1–5 Likert-type scale. In the scale, 1 corresponded to 'I agree completely' ['Ich stimme vollständig zu'] and 5 corresponded to 'I do not agree at all' ['Ich stimme überhaupt nicht zu']. Participants were instructed to focus on the extended-assistance condition when considering their responses. The experimenter read the statements out

[19]The author discussed the grouping with another researcher until agreement was found. The classification and results reported are not the result of a controlled experimental procedure and should be interpreted with respect to that.

and noted down the number that the participant gave as response. See Appendix B.6, for the German questionnaire used in the experiment.

Table 3.14 shows translations of the statements, the mean response, and the corresponding standard deviation. As the table shows, participants made a moderately positive judgement of how well the haptic presentation of the map was understandable (statement 1). They considered the idea to support the exploration of tactile maps with verbal assistance in general helpful (statement 2).

Table 3.14: Evaluation of VAVETaMs and the recognition task: statements and mean agreement (1 corresponds to 'I agree completely' and 5 to 'I do not agree at all').

Statement		**Mean**	**SD**
Number	**Translation**	**Agreem.**	
1	The haptic presentation helps to understand the map.	2.00	.82
2	Giving verbal assisting utterances for tactile maps is helpful.	1.14	.38
3	The actually given utterances were understandable.	1.43	.54
4	I would have liked to get more assisting utterances.	3.14	1.35
5	The utterances were always given at a good time point.	2.57	.54
6	The assisting utterances for the streets were as helpful as for the tree and the buildings.	1.57	.79
7	The task to put the map together in the jigsaw puzzle was understandable.	1.86	1.46

Furthermore, the utterances that they actually had received in the experiment were judged as understandable (statement 3). After indicating agreement with statement 3, participants were asked whether they had received assisting utterances they did not find helpful in the experiment. No participant indicated that any of the assisting utterances was superfluous. Yet, two participants made suggestions for improvements of utterances. One concerned the formulation of the identification-class messages. The participant stated that just verbalizing the name without saying 'that is ...' ['das ist ...'] in the beginning of each phrase would be more efficient. Another participant stated that the assisting utterances that were given are suitable to gain an overview about an environment, but for actual orientation [original German phrase: 'konkrete Orientierung'], more information would be helpful.

Statement 4 was included to assess whether participants wished for more assisting utterances. Their mean neutral answers to this statement do not suggest that they wish for more assisting utterances. If participants indicated agreement to this statement, they were asked whether they can give practical examples for utterances they think would be helpful. Answering this question, one participant stated that he or she would have liked more information where map objects [original German phrase: 'Teile auf der Karte'] were located on the map (e.g., in the left part of the map). Two participants remarked that whether they wish to receive more assisting utterances depends on the goal of map use or the position on the map in combination with the goal of map use. One of them stated that whether additional verbal information is useful depends on the goal of learning the map [original German phrase: 'Hängt vom konkreten Wunsch ab']. Another participant stated that just when being close to the goal location of a possible navigation, more verbal assistance possibly is helpful, otherwise not. In addition, one participant said that a mode in which sidewalks and public transport lines and stops are shown would be useful. Two participants remarked that they would have liked more information about where on the map they are (with their hand movements). One of those stated additionally that he or she would have liked information whether he or she has explored a part of a map before.

The answers to statement 5 indicate that the assisting utterances were not always started in a timely manner. This was due to the large number of buttons to be considered by the experimenter in the Wizard-of-Oz-like experimental scenario in combination with sometimes fast exploration movements of the participants. Participants judged the assisting utterances for the street network as helpful as those for the potential landmarks (statement 6). Furthermore, on average, they found the tactile recognition task understandable (statement 7).

3.7 Experiment 3: Towards an Explanation for the Difference in Performance

3.7.1 Introduction

Two experiments were discussed previously. In these experiments, participants learned virtual tactile maps under two assistance conditions. The purpose of these experiments was to test if users of a virtual haptic interface benefit from extended verbal assistance. Experiment 1 shows that blindfolded sighted people only benefit from extended verbal assistance in their knowledge about potential landmarks, not about tracks (see Section 3.4). In contrast, Experiment 2 shows that visually impaired people benefit in their spatial knowledge about tracks and about potential landmarks when they receive extended assistance (see Section 3.6). Initially, the sub-optimal timely behavior of the Wizard-of-Oz-like setting was considered to be the cause for this result. As participants in Experiment 2 used the maps in the same setting this explanation can be ruled out. Yet, various other explanations are possible:

1. *Attentional focus effect*: For sighted people, potential landmarks such as buildings might be more important in later navigation than for visually impaired people. Sighted people can perceive distant cues much easier and in much larger distance than visually impaired people. Moreover, visually impaired people reported that learning about the track structure is very important for visually impaired navigation in the interviews following Experiment 2. One participant specifically mentioned that information about dead ends (which was included in the extended-assistance condition, but not in the simple-assistance condition) was particularly helpful. For

instance, for visually impaired people, it is demanding to figure out whether a track is a dead end while navigating in the environment. A visually impaired traveller cannot see the end of the street that a sighted traveller potentially can see. In addition, visually impaired people cannot perceive road signs that usually indicate that a street is a dead end. Consequently, it is possible that blindfolded sighted people focus more on potential landmarks when they try to learn the map, whereas blind and visually impaired people focus more on the track structure. Therefore, blindfolded sighted participants benefit only with respect to their knowledge about potential landmarks and visually impaired people benefit additionally with respect to their knowledge about the track structure.

2. *The feeling of completely knowing the map*: Blindfolded sighted people are less experienced with haptic explorations of objects than visually impaired people. Therefore, it is possible that sighted people think they fully know the track structure but have built up an imprecise or partially wrong model of it. They focus on landmark information because they think they fully know the track structure early in their exploration. In contrast, visually impaired people are experienced with haptic explorations. Hence, they detect when conflicts between their spatial knowledge and the track structure on the map occur. Therefore, they continue to learn about the track structure.

3. *Problems with representationally multi-modal integration by sighted participants*: Another hypothesis that can explain the difference between the two groups is that visually impaired participants are more used to integrate spatial language and haptic perceptions in a coherent spatial mental model. Possibly, integrating information about a complex structure such as the track structure on the experimental maps is more complex than integration of information about possibly point-like conceptualized potential landmarks and visually impaired people are better at this task.[20]

Experiment 3 was conducted to gain an insight in the cause of the difference in performance regarding the learning of the track structure. For this purpose, Experiment 1 was modified and repeated. In contrast to Experiment 1, only tracks were present on the maps. The same maps were used, but without potential landmarks. An additional map with comparable complexity was created because a third condition (discussed in the following) was included. The tests of spatial knowledge that were used in Experiment 1 were adapted to fit to the street-only maps. The tests were (1) a verbal test (configuration questions) about coarse-grained spatial facts on the map; (2) sketch mapping; and (3) a recognition task, a task similar to a jigsaw puzzle in which participants were asked to identify correct map parts among multiple options. Additionally, we asked participants for a self-evaluation of their knowledge of the track structure. Moreover, a third

[20] I thank Mark May for the hint that the track structure might be more complex to conceptualize than the location of potential landmarks and, accordingly, it might be more complicated to integrate verbal information about the track structure than about potential landmarks.

learning condition without any assistance was included in the experimental procedure to test whether verbal information helps at all. This condition is in the following called '*no-assistance condition*'.

As described, there were only tracks (and no landmarks) on the experimental maps. If sighted people are unable to integrate the verbal information about tracks that is given in the extended-assistance condition (hypothesis 3), their spatial knowledge should not increase when they learn the maps in this condition in comparison to the simple-assistance condition. In the case hypothesis 2 is true, blindfolded sighted people have the feeling they completely know the maps. This should be reflected in their self-evaluations: Self-evaluations should indicate that participants know the maps after they learned them with simple assistance and should not increase after learning the map with extended assistance. If the effect is due to differences in the attentional focus of the groups of participants (hypothesis 1), blindfolded sighted people should benefit from extended assistance in their knowledge about the track structure when only tracks are on the maps.

3.7.2 Method

Participants

Fourteen participants took part in the experiment. All were able to perform the training procedure. Data of one participant was excluded due to technical problems with the device during the experiment. Consequently, data of 13 participants were evaluated (9 males, mean age: 22.54 years, *SD*: 3.97 years). All participants were compensated by partial course credit or on a monetary basis. They were naïve about the purpose of the experiment and had never seen the experimental setup and were not given any feedback after performing the tests of their knowledge. All participants gave written informed consent. They reported to speak German on a native-speaker level and to have no cognitive disorders. All participants used their self-reported primary hand for the exploration of the virtual tactile map.

Materials and Procedure

Material. In the *no-assistance condition*, participants did not receive any verbal assisting utterances. For the other two conditions, *simple assistance* and *extended assistance*, the set of assisting utterances was as described in Experiment 1 (Section 3.4), but only the information about the track structure was included.

In order to test three conditions in a repeated measures-design, three maps were created. Two of these maps were used in Experiment 1 and 2, for Experiment 3 the potential landmarks were removed. Additionally, a third map was created that had the same number of named streets, crossings, and dead ends. The name sets used for the streets were identical to those used in the previous experiments. There was no need to include a third name set, as the no-assistance condition was performed without verbal assistance. Therefore, each participant only took part two conditions in which the streets had names. Figure 3.21 shows a visualization of the maps used in the experiment.

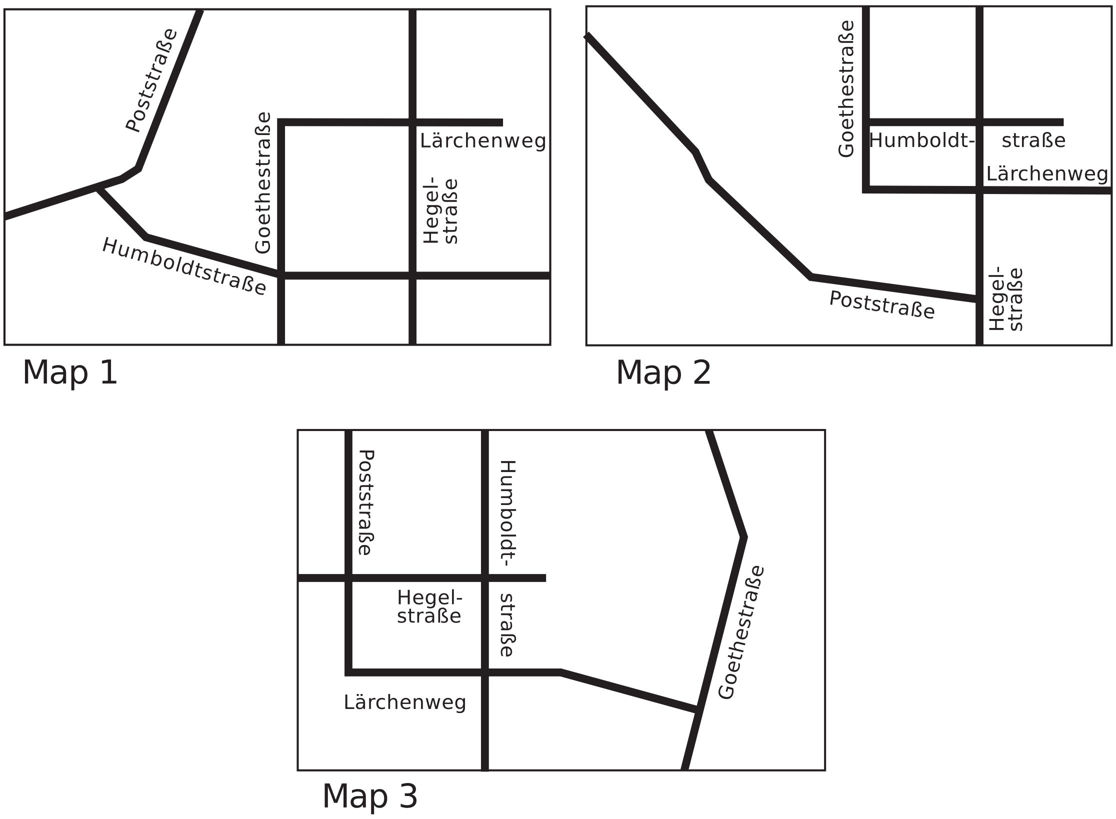


Figure 3.21: Visualization of the maps used in Experiment 3 with the first name set.

Testing Methods. As first task after map exploration, participants were asked to self-evaluate their knowledge about the map they had explored. They were asked to indicate their agreement to two statements on a 5-point Likert-type scale by marking the corresponding box. The leftmost box had the meaning 'is entirely correct' [original German phrase: 'trifft vollkommen zu'], the rightmost box had the meaning 'is not correct at all' ['trifft überhaupt nicht zu']. The two questions were formulated as follows:

1. 'I have precise knowledge of the course of all streets.' ['Ich weiß genau, wie alle Straßen verlaufen.']

2. 'I know the position of all crossings.' ['Ich kenne die Position aller Kreuzungen.']

Second, participants were asked to answer the 20 configuration questions (which were adapted to the track-only maps). The questions that involved landmarks in Experiment 1 and Experiment 2 were removed from the set of questions. Ten questions that were only about the track structure were added. The questions were posed by the experimenter in an individual random order. Table 3.15 shows translations of the questions that participants were asked to answer when they had explored Map 1 with the first name set.

As three maps were used in this experiment, three sets of configuration questions were created. Each question always asked for a comparable spatial fact among the sets. For example, the translation of question 1[21] for Map 2 and the first name set was: 'Do Lärchenweg and Hegelstraße intersect?'. Question 1 for Map 3 with the first name set was: 'Do Poststraße and Hegelstraße intersect?'. As this example shows, question 1 asked about an intersection and was correctly answered with 'yes' for all three maps. Accordingly, the remaining 19 questions for each of the three maps were developed. Since for each map two name sets were used, this procedure resulted in a total of six sets of questions (three maps with two name sets each). The original German sets of questions are shown in Appendix B.5.

As in Experiment 1, participants sketched the map on a sheet of paper. The frame defining the dimensions of the map was printed on the paper for sketching. All sketches were evaluated by the researcher and an independent rater. The rating was performed with regard to which degree the sketch resembles the original map concerning the course of tracks, their parallelism, and the intersections they form. The rating was performed on a 5-point Likert-type scale. A rating of 1 is associated with 'does not reflect the original' and 5 is associated with 'reflects the original precisely'.

The fourth task to test participants' spatial knowledge was a track-only adaption of the recognition task. Again, visualizations of the maps were split in quadrants and for each of the quadrants six optional parts were given. Five parts showed one or two mistakes and one part was correct. Each of the parts fitted to each of the other parts; that is, there were no parts that included tracks that did not connect with the tracks on the

[21] The questions are numbered for the purpose of reference in this description. The numeration does not reflect the order in which they were posed. As described, in the experiment, they were posed in a predetermined, individual random order.

Table 3.15: The configuration questions used in Experiment 3 for Map 1 and the first name set.

Question number	**Question (Translation)**	**Correct Answer**
1	Do Goethestraße and Humboldtstraße intersect?	Yes
2	Are Lärchenweg and Humboldtstraße partially or fully parallel?	Yes
3	Is Hegelstraße upwards restricted by the map frame?	Yes
4	Do Humboldtstraße and Hegelstraße intersect?	Yes
5	Is Humboldtstraße rightwards restricted by the map frame?	Yes
6	Do Humboldtstraße and Poststraße form a T-intersection?	Yes
7	Are Goethestraße and Hegelstraße partially or fully parallel?	Yes
8	Does Lärchenweg form a dead end to the right and is, consequently, not restricted by the map frame?	Yes
9	Do Hegelstraße and Lärchenweg intersect?	Yes
10	Do Goethestraße and Lärchenweg meet?	Yes
11	Do Goethestraße and Hegelstraße intersect?	No
12	Are Lärchenweg and Hegelstraße partially or fully parallel?	No
13	Is Humboldtstraße leftwards restricted by the map frame?	No
14	Do Lärchenweg and Humboldtstraße intersect?	No
15	Is Lärchenweg rightwards restricted by the map frame?	No
16	Do Poststraße and Hegelstraße form a T-intersection?	No
17	Are Humboldtstraße and Goethestraße partially or fully parallel?	No
18	Does Goethestraße upwards form a dead end and is, consequently, not restricted by the map frame?	No
19	Do Poststraße and Lärchenweg intersect?	No
20	Do Humboldtstraße and Lärchenweg meet?	No

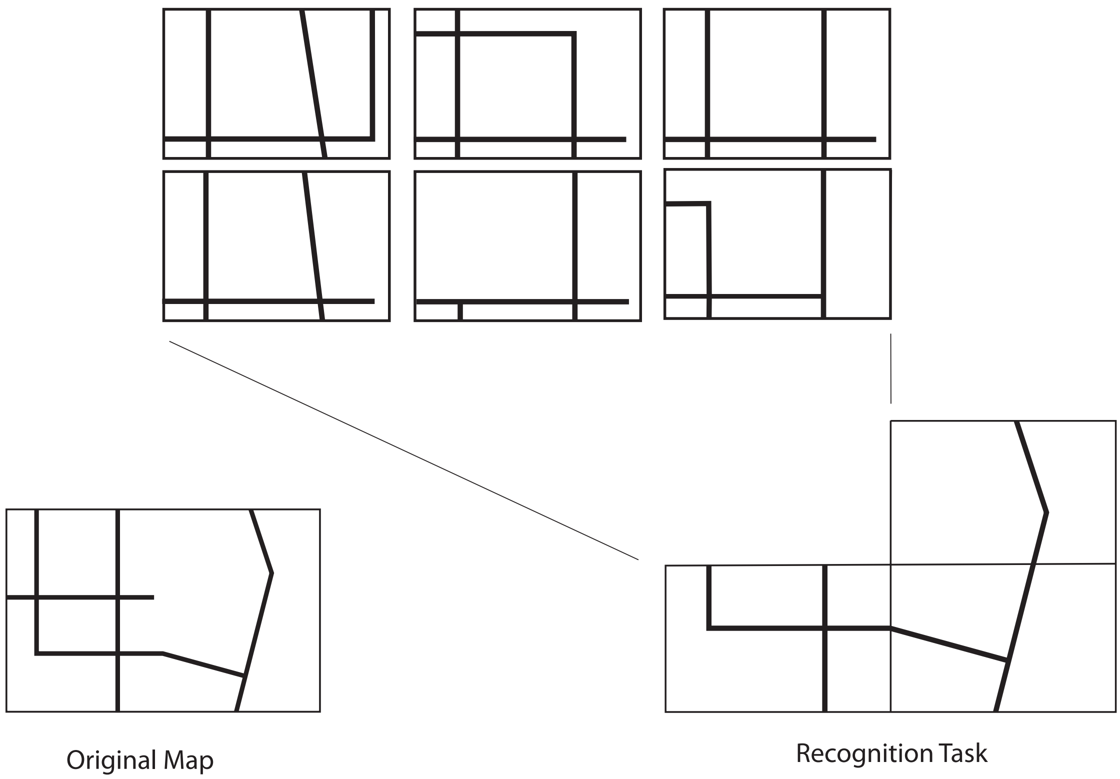


Figure 3.22: The track-only adaption of the recognition task. The example shows the parts for the upper-left part of Map 3. Choice options were given for the other quadrants as well.

other parts. Therefore, the solution for each quadrant was independent of the solution of other quadrants. The goal position and the orientation of parts was given. Consequently, participants just had to choose which part is the correct one for each of the quadrants and did not have to make a decision about its orientation or which part belonged where. Figure 3.22 shows a visualization of the optional parts for the upper left quadrant of the map visualization. As can be seen in the figure, different types of error were included in the track layout on the mistakes—for example, the course of the streets was wrong, streets or part of streets were missing or added.

Procedure. Initially, a design based on latin squares was developed to control for order effects (a latin-squares design organizes the order of conditions in a repeated-measures experiments in a way that allows for controlling order effects; see Bradley, 1958; Field & Hole, 2003, for a discussions of latin squares designs). Due to the complexity of the combinations (controlling for order of conditions and combinations of maps, name sets, and conditions was desired), two latin-square designs were combined. This resulted in nine combinations of the order of conditions, the order of maps, and the order of name sets. These combinations are shown in Table 3.16.

Table 3.16: The nine possible combinations of the order of conditions, the map used, and the name set used.

Combi-nation	**Order of Conditions (Map Used, Name Set Used)** ex = extended assistance, si = simple assistance, no = no assistance
1	ex (Map 3, Name Set 2) → si (Map 2, Name Set 1) → no (Map 1, –)
2	si (Map 3, Name Set 2) → no (Map 2, –) → ex (Map 1, Name Set 1)
3	no (Map 3, –) → ex (Map 2, Name Set 2) → si (Map 1, Name Set 1)
4	ex (Map 2, Name Set 1) → si (Map 1, Name Set 2) → no (Map 3, –)
5	si (Map 2, Name Set 1) → no (Map 1, –) → ex (Map 3, Name Set 2)
6	no (Map 2, –) → ex (Map 1, Name Set 1) → si (Map 3, Name Set 2)
7	ex (Map 1, Name Set 2) → si (Map 3, Name Set 1) → no (Map 2, –)
8	si (Map 1, Name Set 2) → no (Map 3, –) → ex (Map 2, Name Set 1)
9	no (Map 1, –) → ex (Map 3, Name Set 2) → si (Map 2, Name Set 1)

Participants were assigned to a group in the following manner: Each participant was assigned randomly to one of the nine combinations of order of conditions, combination of map and condition, and combination of condition and name sets of a latin-squares design. Each participant was assigned to a combination that no participant had performed yet. After nine participants, each of the possible experimental combinations of the latin-squares design was used for data collection with one participant. Therefore, for the following participants we proceeded in the same way—that is, we randomly assigned them to one of the possible experimental combinations. With the data of 13 participants evaluated, not all experimental combinations of the latin-squares design were assigned to two participants, some were only assigned to one participant. As described, this assignment was randomized.

As in Experiment 1 and 2, participants were trained prior to the experimental procedure. In this training, they learned the interaction with verbally assisted maps with a map different from the three training maps. On the training maps, like on the experimental maps (Map 1–3), no landmarks were present. Like in the previous experiments, after participants knew how to interact with the system, the training success was tested by asking them to follow a structure similar to a labyrinth. During this task, the experimenter gave assisting utterances that informed them about where to move. Participants had to follow the tracks from beginning to end without ever leaving the track structure, like in Experiment 1 and 2. After the training procedure, participants read written instructions that included information about the tasks they had to solve after the learning conditions.

Participants were instructed to learn the map such that they know each street on the map (see Appendix B.3, for the instruction). In each condition, participants were allowed to explore the map for four minutes. The exploration time was shorter than in Experiment 1 and 2 to avoid ceiling effects (because the potential landmarks were left out, there were less objects on the maps). After that, they were asked to complete the tasks described earlier in this section. After participants learned maps in the simple-assistance and extended-assistance condition, all tasks were applicable. In the no-assistance condition, participants did not receive any assisting utterances. Therefore, the configuration-questions task was not applicable in this condition and was left out when participants learned the map in this condition. Between each of the learning conditions, a short break was made. The experiment took about two hours per participant including the training session.

3.7.3 Results

Self Evaluation. The first task participants solved after they completed the exploration was to self-evaluate how well they knew the map. Therefore, they were asked to indicate agreement to two statements on a Likert-type scale, as discussed. The first statement was about how well they think they knew the course of tracks. The second statement was about whether they knew the position of all the crossings on the map. For each of the statements, a repeated-measures ANOVA was calculated.

The agreement data for the first statement show a significant effect of assistance condition ($F(2,24) = 27.94$, $p < .001$, $\omega = .65$) with a large effect size indicating a substantial effect (according to Field, 2005). There was a significant linear trend ($F(1,12) = 103.143$, $p < .001$), indicating that the more assistance participants received, the more they felt confident to have learned the course of the streets on the map well.

The data for the second statement show a similar pattern. They show a significant effect of assistance-condition ($F(2,24) = 18.58$, $p < .001$, $\omega = .58$) with a large effect size. There is a significant linear trend ($F(1,12) = 70.59$, $p < .001$), indicating that the more assistance participants received, the more they felt confident to have learned the position of the crossings on the map well. Both self-evaluations correlate significantly and to a large degree ($r = .718$, $p < .01$ (two-sided)).

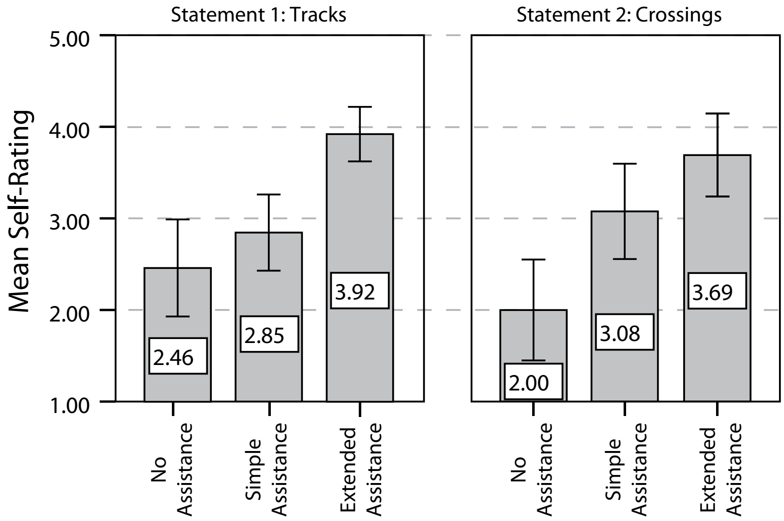


Figure 3.23: Mean self-ratings after map learning with the different assistance conditions. Error bars represent the 95 % confidence interval of the mean.

Figure 3.23 shows the mean self-evaluations for the different conditions for the two questions. As the figure shows, participants rated their knowledge best after they explored the map in the extended-assistance condition. The self-evaluations after the no-assistance conditions were worst and the self-evaluations after the simple-assistance condition in between the other two.

This finding indicates that participants have the feeling of knowing the virtual tactile map best when they learned the map receiving assistance in the style of the extended-assistance condition. Participants' self-evaluations after learning the map with the simple-assistance condition are higher than after learning the map with the no-assistance condition.

Configuration Questions. As described, the configuration-questions task was applied after the simple-assistance and the extended-assistance condition, but not after the no-assistance condition. Participants were asked to answer 20 questions about the map. Figure 3.24 shows the mean number of correct answers (simple assistance: $M = 13.23$, $SE = 1.03$; extended assistance: $M = 15.38$, $SE = .84$). Participants were able to answer more questions correctly after they learned the map with extended assistance. A paired t-test shows the significance of this finding ($t(12) = 2.24$, $p < .05$, $r = .54$). The large effect size indicates a substantial effect (Cohen, 1992). This indicates that participants' knowledge of the track structure is improved after they learned the map in the extended-assistance condition compared to the simple-assistance condition.

Sketch Mapping. As done for the evaluation of the sketch maps in Experiment 1 and Experiment 2, the sketch maps were evaluated on a 1–5 Likert-type scale. Two inde-

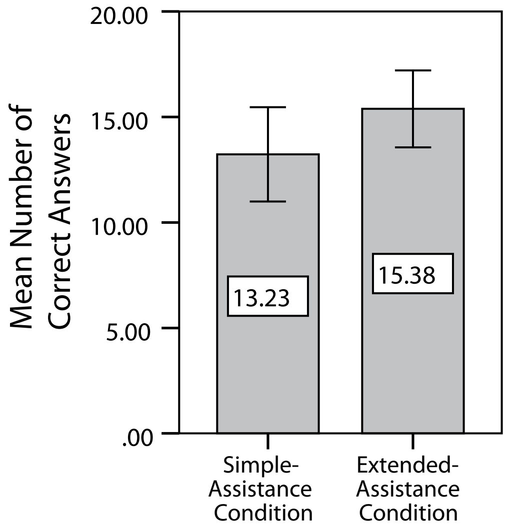


Figure 3.24: Mean number of correctly answered configuration questions. Error bars represent the 95 % confidence interval of the mean.

pendent raters evaluated the sketch maps with respect to how well they reflect the track structure in comparison to a visualization of the original map. The raters were blind to the learning condition after which the maps had been produced. Furthermore, they were blind to which participant made the sketch. To control for inter-rater reliability, the intraclass correlation coefficient was calculated (Shrout & Fleiss, 1979). The raters' agreement is fair (*ICC(3,1)* = .813).

Figure 3.25 shows the mean ratings of the sketches. As the figure shows, the sketches that participants produced after the simple-assistance ($M = 2.65$, $SE = .27$) and extended-assistance condition ($M = 2.65$, $SE = .29$) were, in the mean, rated equally well. The sketches that were produced after the no-assistance condition were rated less well.

A repeated-measures ANOVA was calculated to test whether assistance condition significantly affected the quality of the sketches. The analysis showed that the effect of assistance condition is significant ($F(2, 24) = 6.28$, $p < .01$). Contrasts comparing the other conditions to simple-assistance revealed a significant difference between no-assistance ($M = 1.73$, $SE = .24$) and simple-assistance condition ($F(1, 12) = 10.70$, $p < .01$, $r = .69$). Between mean sketch ratings in the simple-assistance condition and those that were produced after learning in the extended-assistance condition, there was no significant difference ($F(1, 12) = 0$, $p = 1.0$, $r = 0$).

Altogether, the data indicate that participants' sketch maps were better when they learned the map with verbal assistance, but it did not matter which style of assistance (extended assistance or simple assistance) they received.

Recognition Task. To evaluate the data of the recognition task, they were transformed in the following way: As overall eight mistakes were possible, we subtracted the number

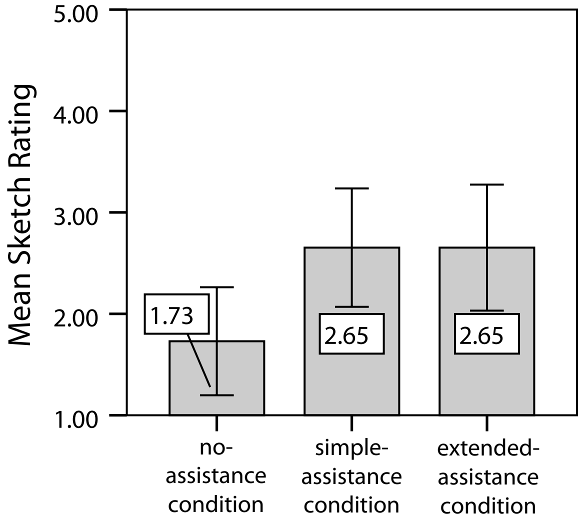


Figure 3.25: Mean rating of the sketch maps. Error bars represent the 95 % confidence interval of the mean.

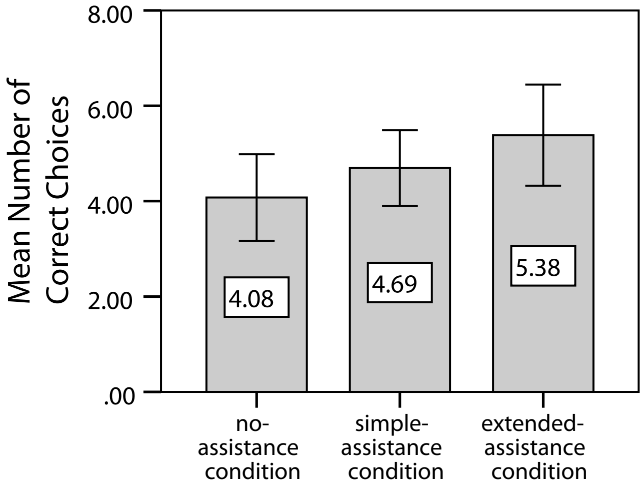


Figure 3.26: Mean number of correct choices in the recognition task. Error bars represent the 95 % confidence interval of the mean.

of mistakes made by the participants from eight. This procedure leads to a measure of correct choices in the task. The mean number of correct choices after learning the map with the different assistance conditions is shown in Figure 3.26. In average, participants were able to make the best choices when they learned the map in the extended-assistance condition ($M = 5.38$, $SE = .49$). The second best learning condition is simple assistance ($M = 4.69$, $SE = .36$) and the worst is no assistance ($M = 4.08$, $SE = .42$). However, a repeated-measures ANOVA revealed that the effect is a non-significant trend ($F(2, 24) = 3.22$, $p = .058$, $\omega = .23$).

The result is not significant. Yet, it supports the interpretation that learning the map with simple assistance is more effective than without assistance. This is in line with the results from the configuration-question data: participants' knowledge acquisition was more effective when they received assisting utterances in the style of the extended-assistance condition.

3.7.4 Interview and System Evaluation

Like in Experiment 2, after the experimental procedure was concluded, participants were interviewed and indicated their agreement to a number of statements. Participants were asked to remember back to the extended-assistance condition (see Appendix B.10 for the original interview guideline). As discussed in Section 3.6.5, two strategies could be identified in the descriptions participants gave in the interviews: (1) the fix-object strategy, following which participants focus on a salient part of the map and systematically explore the objects starting from this part, and (2) the overview strategy, following which participants first try to gain an overview about all map objects. On the basis of the interviews, it was evaluated whether participants used one of the two strategies. Six participants' interviews indicated that they were using a fix-object strategy. Likewise, six other participants' interviews indicated that they were using an overview strategy. One participant's answers did not allow for a clear classification. Like the results of Experiment 2, the results of Experiment 3 do not show an effect of strategy use.

As in Experiment 2, after they were interviewed, participants indicated agreement to statements about the system. See Appendix B.9 for the original German questionnaire. Again, participants gave their response on a 1–5 Likert-type scale in which 1 corresponded to 'I agree completely' ['Ich simme vollständig zu'] and 5 to 'I do not agree at all' ['Ich stimme überhaupt nicht zu']. Table 3.17 lists translations of the statements and the mean agreement with them. Statements 1, 2, 3, 6, and 7 were formulated like in Experiment 2 (there, the corresponding statements were 1, 2, 3, 4, and 5). Additional statements were included to control whether participants were able to understand the meaning of the assisting utterances (statement 4) and considered the statements helpful (statement 5).

All statements that were used in both experiments were similarly responded in both experiments (maximum difference of the mean: .35), even though whether landmarks where included on the maps and the group of participants differed between the two experiments. As discussed, sighted participants in Experiment 3 explored maps without potential landmarks whereas in Experiment 2, potential landmarks were included on the experimental maps explored by visually impaired participants.

Table 3.17: Evaluation of VAVETaMs with only tracks: Statements and mean agreement (1 corresponds to 'I agree completely' and 5 to 'I do not agree at all').

Statement		**Mean Agreem.**	**SD**
Number	**Translation**		
1	The haptic presentation helps to understand the map.	2.00	.60
2	Giving assisting utterances for tactile maps is helpful.	1.17	.39
3	The actually given utterances were understandable.	1.08	.29
4	I always knew what was meant with the verbal utterances.	1.42	.67
5	The actually given utterances were helpful.	1.42	.67
6	I would have liked to get more assisting utterances.	3.25	1.42
7	The utterances were always given at a good time point.	2.50	.91

Table 3.18: Additional statements about map use: Mean agreement (1 corresponds to 'I agree completely' and 5 to 'I do not agree at all').

Statement		**Mean Agreem.**	**SD**
Number	**Translation**		
8	I am able to orient myself well in a city new to me using a city map.	1.50	.67
9	Even with a city map, I sometimes loose orientation in a city new to me.	3.83	1.03

The high agreement to statement 2 (about verbal assistance in general) indicates that the approach to assist a tactile-map user verbally was considered helpful by participants. Furthermore, the high agreements to statements 3, 4, and 5 indicate that the assisting utterances were formulated in a well-formed manner for both groups of participants. In line with the results from Experiment 2, the medium agreement to statement 6 shows that participants did not want to receive more assisting utterances. Yet, the results show a high variance in the answers to this statement. The medium agreement with statement 7 indicates that the utterances were not always started in a perfect timely manner in the Wizard-of-Oz-like experimental setting.

Table 3.18 shows statements 8 and 9, which were included to test whether the facility of general, visual map use as reported by the participants has an effect on the facility with which they acquire knowledge using VAVETaMs. The mean agreements to the two statements correlate significantly and negatively[22] ($r = -.79$, $p < .01$). The mean agreement to the two statements correlates significantly with the number of correct answers that participants could give in the configuration questions after learning the map with extended assistance (statement 8: $r = -.58$, $p < .05$; statement 9: $r = .58$, $p < .05$). Participants who indicated to be able to orient themselves with maps in novel environments were able to answer more configuration questions correctly. This indicates that the use of VAVETaMs demands similar abilities like using a visual map for orientation.

3.7.5 Discussion

Two previously reported experiments, Experiment 1 and 2, showed that sighted and visually impaired participants perform differently with respect to how extended verbal assistance supports spatial knowledge acquisition. Sighted participants did only improve in their knowledge about potential landmarks. They did not improve regarding their knowledge about the track structure. In contrast, visually impaired participants did improve in both respects. The experiment reported was designed to test three hypotheses for explaining this difference. Hypothesis 1 was that participants' attentional focus differs depending on whether they are sighted or not because different kinds of information are important in navigation. Hypothesis 2 was that sighted individuals have the feeling they completely know the track layout when, in fact they do not. They do not realize that the haptic presentation differs from their spatial mental model. According to hypothesis 3, sighted individuals are, in contrast to visually impaired individuals, not able to integrate verbal and haptic information.

The data show that hypothesis 2 can be ruled out: Participants mean self-evaluations indicate that they do not have the feeling of completely knowing the track structure after they have learned the map with simple assistance. Therefore, it is unlikely that they do not acquire additional knowledge about the track structure when they receive extended assistance due to wrongly assuming that they already completely know the track structure with simple assistance. The data show a significant linear trend from no-assistance condition to extended-assistance condition. Map users think that they

[22]The negative correlation results from the negative formulation of statement 9.

know more about the map when they receive more assisting utterances. This indicates that participants considered the extended-assistance condition the most helpful style of assisting utterances.

In addition to self-evaluations, adapted versions of the testing methods used in Experiments 1 and 2 were applied. These methods were (1) configuration questions, (2) sketch mapping, and (3) a recognition task. The results of these tasks do not allow for a clear interpretation. With the exception of the data of the sketch mapping task, the data of all tasks support that sighted participants improve in their knowledge about tracks when they have learned the maps in the extended-assistance condition compared to the simple-assistance condition. The configuration-questions task shows this significantly. Overall, this suggests the interpretation that participants generally benefit from the additional verbal information given in the extended-assistance condition. However, more empirical validation is needed, especially to explain the virtually equal performance in the sketch-mapping task after learning the map with simple and with extended assistance.

The different outcomes of the testing methods can possibly be explained as follows: The configuration questions and the recognition task benefit to a larger degree from propositional knowledge that was explicitly stated in the assisting utterances (e.g., whether a track is a dead end). Purely propositional information is sufficient to answer the configuration questions correctly. Additionally, it helps to solve the recognition task, at least by applying an exclusion procedure. For example, if participants know that a track in the upper left part of Map 3 forms a dead end to the right, they can exclude two incorrect options for the upper left part (see Figure 3.22, Section 3.7.2). They can do so without having integrated that information in a coherent spatial mental model containing an analog representation for this track and without knowing something else about the track (such as its course). In contrast, in sketch mapping, additional knowledge about the course of the track is needed to sketch it correctly and purely propositional knowledge—such as knowing that a track in the upper left part of the map forms a dead end to the right—is not sufficient to solve the task.

Possibly, the information that is given verbally is accessible (e.g., in a verbatim manner) in testing, but not integrated with the haptic percepts into a coherent spatial mental model. In this case, it is likely that participants improve in tests that can be solved solely with propositional knowledge. This argument is supported by empirical evidence that (sighted) individuals are able to code or retrieve map knowledge both in analog and in propositional fashion. Which representation is prominent differs among (sighted) individuals (MacEachren, 1992, 2004). Given differences in coding preferences among sighted individuals, differences between sighted and visually impaired individuals are likely. Research has shown differences between sighted and blind people in the processing of spatial information from different sensory modalities (for an overview, see Loomis et al., 2012). Struiksma (2011) has investigated the processing of spatial sentences and non-representational haptic stimuli by congenitally blind, visually impaired, and sighted people. While some sentences lead to a highly similar result for all groups, other sentences did not (dependent on the reference systems used). Struiksma speculates that differences can be explained by increased experience with haptic object handling by visually impaired people. Whether processing differences with respect to representationally multi-modal information displays, such as VAVETaMs, exist, has not systematically been investigated in the literature.

A possible counterargument against the hypothesis that spatial knowledge is not integrated by sighted participants is that the results of Experiment 1 do not show differences between the tasks. However, it is possible that potential landmarks are point-like conceptualized (Tversky, 1993). Therefore, verbal information regarding them is easier to integrate than information about the track structure, which is more complex. Furthermore, in Experiment 3 participants were given less time to explore the maps. Possibly, a longer exploration time leads to a higher degree of integration of spatial knowledge. Moreover, an *additional* effect of attentional focus can explain the difference between the results of sighted and visually impaired people. When forced (as in Experiment 3) to focus their attention on the track structure, sighted people are able to use extended assistance to acquire knowledge about tracks. Still, they do not integrate this knowledge completely.

Actually, when the results of the tactile sketch mapping in Experiment 2 are reconsidered, the very similar results of the ratings of the track structure (simple-assistance condition: $M = 2.29$, extended-assistance condition: $M = 2.50$) could indicate that improvement only showed in the configuration questions. Still, the difference between the landmark ratings is small, as well (simple-assistance condition: $M = 1.64$, extended-assistance condition: $M = 2.07$). As discussed, the sensivity of the tactile adaptions of sketch mapping is questionable (see Sections 3.6.3 and 3.6.4). Altogether, it is unclear whether the difference in performance of visually impaired and sighted people stems from differences in attentional focus, from differences in cognitive abilities, or from a combination of both. The results encourage further research on how sighted and visually impaired people integrate spatial knowledge from representationally multi-modal sources.

In Experiment 3, the no-assistance condition was included in addition to the conditions of Experiments 1 and 2 to test whether assisting utterances help at all. The recognition task showed a non-significant trend and sketch mapping a significant difference (configuration questions were not applicable). This finding is especially noteworthy as both tests of spatial knowledge were tasks that rely purely on the spatial layout of the tracks and not on knowledge about the names of tracks. Still, participants perform better in tasks concerning the spatial layout when they are informed about the names of tracks. Possibly, knowing the names of the tracks enables to recognize the identity of them faster and more accurate and prevents to get lost (not knowing the own position in relation to known tracks) during exploration of the maps (getting lost is a phenomenon occurring in virtual haptic environments also described by Colwell et al., 1998a; Colwell, Petrie, Kornbrot, Hardwick, & Furner, 1998b).

As discussed in Section 3.7.4, following the experiments, participants were asked to evaluate the system by indicating their agreement to statements on a Likert-type scale. Participants indicated agreement to the same statements as used in Experiment 2. The mean response to the statements in Experiments 2 and 3 was very similar. Participants evaluated both assisting utterances for tactile maps in general and the assisting utterances that were actually given in the extended-assistance condition as helpful and understandable. The data indicate that the timing of assisting utterances in the Wizard-of-Oz-like setting was not always ideal. In addition, the data indicate that participants in average did not miss specific utterances or wish for additional assisting utterances in the extended-assistance condition.

Altogether, Experiment 3 shows that verbal assistance on track-only maps facilitates spatial knowledge acquisition in comparison to giving no assistance. The differences in the outcomes of the testing methods do not allow to conclude whether hypothesis 1 (attentional focus), hypothesis 3 (integration of information), or a combination of both explain the differences. Yet, hypothesis 2 (feeling of knowing the track structure) can be ruled out. Additional research is needed both to clarify the differences among the outcomes and to fully explain the difference between the performance of sighted and visually impaired participants.

3.8 Chapter Summary

Knowledge acquisition of virtual tactile maps for visually impaired people can be facilitated by verbal assistance including information about the spatial relations and configurations of objects on the map. The set of assisting utterances that is suggested (extended assistance) is based on a corpus study in which human assistants were asked to help a blindfolded map explorer to acquire knowledge from a virtual tactile map.

Three experiments were conducted to test the effectiveness of VAVETaMs. In these experiments, extended assistance was compared to other kinds of verbal assistance including less information. As recommended in the literature, three mutually supportive tests of spatial knowledge were used: (1) a verbal task (called 'configuration questions'), (2) sketch mapping, and (3) a recognition task (similar to a jigsaw puzzle).

Experiment 1 was conducted with blindfolded sighted participants. Regarding overall performance, they acquired significantly more knowledge when they learned virtual tactile maps with extended assistance in comparison to a condition with less verbal information (simple assistance). A subsequent analysis revealed that this knowledge increase was only based on additional knowledge about potential landmarks. There was no increase in knowledge about tracks.

With Experiment 2, we basically repeated Experiment 1 with visually impaired participants, which are the potential user group of VAVETaMs. In contrast to blindfolded sighted individuals, visually impaired individuals benefit in both their knowledge about the configuration of potential landmarks and their knowledge about the track structure.

To explain the difference between the groups of participants, Experiment 3 was conducted. In Experiment 3, blindfolded sighted participants learned virtual tactile maps that only included tracks and did not include potential landmarks. Additionally to the simple-assistance and extended-assistance condition, an assistance condition without any verbal assistance (no-assistance condition) was included. The results show that simple assistance facilitates the acquisition of knowledge of the track structure in comparison to no assistance. Aside from this finding, the results of Experiment 3 do not allow for a clear interpretation with respect to the effectiveness of extended assistance for track-only maps for sighted people or the difference between the user groups. Yet, the hypothesis that sighted participants have the feeling to know the track structure completely and do not realize conflicts with their spatial mental models can be ruled out.

Following the experimental procedure of Experiments 2 and 3, participants were interviewed and and asked to evaluate VAVETaMs on Likert-type scales. Both in the interview and in the scale-based evaluation, sighted as well as visually impaired participants indicated that they found the assisting utterances that were actually given helpful and understandable. Participants generally indicated that augmenting virtual tactile maps by assisting utterances facilitates knowledge acquisition. Both groups agreed in that they considered all assisting utterances helpful and that nothing in the set of assisting utterances should have been left out. The set of assisting utterances was considered complete by both groups: There were no suggestions to add verbal information, except from when the maps should be used for purposes additional to survey-knowledge acquisition (such as learning a route to be traveled without vision).

In the three experiments, all 50 sighted, blind, and visually impaired participants except one were able to pass the training procedure concluded by a test of the training success. This success rate indicates that in general, VAVETaMs are feasible and the interaction with them is easy to learn. The results of Experiment 2 show that a system realizing VAVETaMs can increase the mobility of visually impaired people by providing an effective means for the communication of survey knowledge. Visually impaired participants benefited significantly from receiving extended assisting utterances.

Chapter 4

Building an Intelligent Interface: Technique[1]

4.1 Introduction

The generation of assisting utterances in a human-inspired manner is a complex task for an artificial system. As discussed in Chapter 2, existing approaches towards audio-tactile maps are based on associating fixed sounds or fixed utterances to map objects (that is, sound playback or synthesis of text is triggered when the map user touches an object). To generate verbal output similar to that which human assistants give, a new approach towards audio-tactile maps has to be developed: Due to the context-dependent variability of information, it is not possible to strictly associate pre-determined natural-language output with a map object.

The approach I present is based on a representation of exploration events that the map explorer performs, while exploring the map with the device. Semantic (i.e., meaningful) exploration events are detected and represented in a manner suitable to serve as input for a natural-language generation component. Assisting utterances for a particular exploration situation are constructed on the basis of a knowledge base storing knowledge suitable for verbalization. More precisely, the following issues have to be addressed to develop an artificial system with the capability of generating situated assisting utterances as proposed and evaluated in Chapter 3:

- The map user's exploratory hand movements have to be analyzed and conceptualized. It is not sufficient to control the verbal output just on the basis of whether an object is touched (as done in current audio-tactile map systems; see Chapter 2, for a discussion).

- It is necessary to keep track of which verbal information the map user already has received. This is especially important to avoid unnatural, unnecessary, and potentially confusing repetitions of utterances. For example, consider a situation in which a map user explores a track A. The map user verbally receives the information that the track A has crossings with the tracks B and C. Then, the map user stops exploring track A for some seconds—for example, to explore a landmark nearby.

[1] Parts of the content of this chapter are based on previously published content (Lohmann et al., 2011; Lohmann, Eichhorn, & Baumann, 2012; Lohmann, Kerzel, & Habel, 2012).

Subsequently, the map user explores this track again. In this situation, other verbal information is more relevant than to repeat that the track A intersects with the tracks B and C.

- Following from the requirement to construct utterances with respect to what has already been said, a strict association of verbal information to exploration events (or to map objects, as in trigger-based scenarios) is too inflexible. It does not allow to take verbalization history into account for the planning of the assisting utterances. The VAVETaM language-generation component needs to be equipped with a knowledge base that enables the construction of suitable utterances in a flexible manner. Which information is selected to construct these utterances depends on the exploration movements the map user currently performs with the device and on which information was given before (as exemplified previously in the last list item).

In this chapter, I discuss a conception of a natural language generation component for the VAVETaM system and I present a prototype implementation of such a component. I focus on the tasks associated with deciding what should be said in an exploration situation. This chapter is structured as follows. In the next section, I discuss related research in natural language generation. Subsequently, I examine the tasks that have to be solved in a system that is capable of the generation of natural language. In Section 4.3, I discuss how users explore virtual tactile maps. The structure of the VAVETaM system (i.e., the proposed components and their interaction) is described in Section 4.4. While the description of other components of the VAVETaM system is essential to illustrate the interaction of the components with the generation component, the research reported is focused on the generation of verbal assistance for virtual tactile maps. Therefore, I do not discuss how the detection of MEPs, an instance of event detection in haptic environments, is realized.[2] In the following sections, the representation of semantic knowledge is discussed (Section 4.5) and the input to the generation component is discussed (Section 4.6). The generation component of the VAVETaM system and a prototype implementation are presented (Section 4.7). We were able to connect an MEP detection component to the generation component discussed in this thesis and, thus, to create a prototype of the VAVETaM system (Lohmann, Kerzel, & Habel, 2012). The relevant results of a user study are discussed in Section 4.8.

4.2 Situated Natural-Language Generation: An Overview

The area of natural language generation is concerned with the automatic generation of natural language in the form of understandable verbal or textual output from non-linguistic data. As mentioned in Chapter 1, natural language generation is a subfield of artificial intelligence and computational linguistics (Reiter & Dale, 2000). In the following, I discuss some systems that generate output for dynamic or interactive domains

[2] See Kerzel & Habel, 2011, for a discussion of an approach of how such an event detection can be realized.

(an overview about natural-language generation systems in general can be found in Reiter & Dale, 2000).

The goal of the VITRA project (Herzog & Wazinski, 1994) is the generation of simultaneous descriptions of dynamic scenes. One of the application domains is the generation of simultaneous descriptions of soccer games by a system called 'SOCCER' (André, Herzog, & Rist, 1988; Herzog & Retz-Schmidt, 1990). The scenes that are described consist of series of pictures in which events are detected. The events that are detected control the generation of utterances. They are detected on the basis of event models, which represent typical events in a domain (André et al., 1988). An event model consists of several sub-events (which are called 'sub-concepts' in the original publication; André et al., 1988). In that respect, the event model is similar to that we propose in Habel et al. (2010). With respect to the fact than an event model is underlying the conceptualization of natural language, the conception of the VITRA project is similar to the approach discussed in this thesis.

Becker, Kilger, Lopez, and Poller (2000) discuss a component which solves the task to produce simultaneous (i.e., 'online') translations. According to Becker and colleagues, determining the content of utterances, which is central to the work I report, is not necessary in this application. Therefore, the solutions developed for this component are of limited relevance for the research reported in this thesis.

The COLLAGEN system was suggested to enable intelligent user interfaces (Rich, Sidner, & Lesh, 2001). The system is used to communicate with the user using speech-recognition and language-generation technology. Similar to the generation component of the VAVETaM system, a component called 'agenda' stores a list of utterances whose articulation would contribute to achieving the discourse goal.

The SCRISP system presented by Garoufi and Koller (2010) generates situated referring expressions in a visual virtual environment. The virtual environment only allows for stepwise interaction. While the content is based on the output of a planner, the SCRIPS system focusses on the generation of appropriate referring expressions and, thus, solves a different task than the task the VAVETaM system is focussed upon.

Several situated generation systems have been suggested in the context of the GIVE 2.5 challenge, which had the purpose to compare the systems (Striegnitz et al., 2011). In the challenge, natural-language systems guide the users through a virtual environment (e.g., Garoufi & Koller, 2011; Dethlefs, 2011). The systems focus mostly on the generation of appropriate referring expressions or on planning how to formulate the utterances. In contrast to the VAVETaM system, the generation situation is different: While in these systems, natural language is used to guide the user to accomplish a given goal (see also Benotti & Bertoa, 2011), in VAVETaMs, natural language augments the virtual tactile map. Therefore, the timing constraints are different in the two scenarios: In the systems developed in the context of the GIVE 2.5 challenge, the system gives instructions, which users must follow after they are stated. In contrast, in VAVETaMs, the interaction has a more simultaneous character. While the users explore the map haptically, they receive assisting utterances augmenting the haptic perception.

Technically, classical architectures for natural-language generation systems organize the tasks to be solved in (at least) two subclasses. First, a semantic representation that is suitable to generate natural language is created from the input. This task is called 'conceptualization' or the 'What to say?' task (the term 'conceptualization' is introduced in Levelt, 1989, see also Habel & Tappe, 1999; see De Smedt, Horacek, & Zock, 1996, for a discussion how conceptualization relates to components used in natural-language generation systems). In conceptualization, the content of the utterances to be produced is determined in accordance with the input to the generation system.

Human speakers conceptualize on different levels when they describe dynamic scenes: "In producing an utterance, mental representations of perceived or conceived states of affairs are first transferred to an intermediate level of propositional representations and are subsequently transformed into grammatical structures." (Habel & Tappe, 1999, p. 2) The idea of different levels of conceptualization is of particular relevance for the VAVETaM system: First, dynamic exploration movements are detected in the stream of movement data that the device outputs. This detection results in a propositional representation of the current exploration situation, which is an intermediate representation needed for conceptualization of assisting utterances. Second, knowledge is selected for verbalization and a representation, which is suitable as input to following steps of assistance generation (formulation, discussed subseqently), is created. In the VAVETaM system, two different components are concerned with these two levels of conceptualization. The generation component, which this thesis is centered on, solves the task to prepare a representation of the semantics of assisting utterances in a manner suitable for the next processing steps. The selection of information to be verbalized is controlled by a propositional representation of the current exploration situation, which is created by another component by segmenting the map user's exploration movements, as detailed in Section 4.4.

After conceptualization, the semantic representation of utterances is transformed into linguistic output. The corresponding task is called 'formulation' or the 'How to say?' task. This task can (but not necessarily has to) be organized in the following subtasks: sentence planning, surface generation, morphology, and formatting (if the output format is text) (Reiter, 1994). When verbal output is the goal of the generation process, additionally, the result of formulation has to be articulated—that is, uttered in a verbal form. In the current implementation of the generation component for the VAVETaM system, a template-based approach is used (Reiter & Dale, 2000; van Deemter, Krahmer, & Theune, 2005, for a discussion of template based approaches). Ready-made text parts (also called 'canned text') are used to construct a surface form from the semantic representations. As Mellish (2000) points out, such an approach can be described as "taking a shortcut" (p. 3) and is often useful and necessary when working applications or other parts of the language-generation process are the goal of research—both is the case for the generation component of the VAVETaM system.

In the VAVETaM generation component preverbal messages (PVMs) are the semantic representations of the utterances to be expressed, as suggested by Levelt (1989). The same approach is made in the InC system (Guhe, Habel, & Tappe, 2000; Guhe, Habel, & Tschander, 2004; Guhe, 2007). Preverbal messages are semantic representations of the informational content of utterances. They are the interface between conceptualization and formalization. They are propositional representations of a special kind (according to Levelt, 1989, representations of the semantics of natural language have to be in a propositional form). To serve as an interface between conceptualization and formulation, a preverbal message must contain information needed for verbalization. For example, it must contain information about the mood of the utterance—that is, whether the utterance is imperative, declarative, or interrogative. Information about the tense of the utterance is important if the goal language has a tense system (Levelt, 1989). However, for the development of the VAVETaM generation, including this information explicitly was not necessary because the utterances are always formulated in present tense and they are always formulated in a declarative mood.

Generally, two approaches towards natural-language generation systems can be distinguished: *cognitively oriented* approaches and *application-oriented* approaches. Cognitively oriented approaches aim to model certain aspects of human behavior. For example, InC aims to computationally model the behavior that humans think while they are speaking (Guhe et al., 2004; Guhe, 2007). Cognitively oriented approaches have the goal to provide insights in the nature of human information processing by developing models that produce comparable output given comparable input. In contrast, application-oriented approaches do not aim for a modeling of human behavior. The aim of application-oriented approaches is to construct systems that work in a manner such that they satisfy the requirements. The systems reported earlier (except for InC, which is cognitively oriented) can be classified as application-oriented approaches. From an application-oriented perspective, it is irrelevant whether humans process information in a similar manner. Depending on the system, it is not useful to model human output in all respects for an application-oriented approach. An example from the map domain can illustrate this: Even if humans confuse 'left' and 'right' when they give assisting utterances to a map user and then correct their utterances, a computational generation system is probably more appropriate if it straightforwardly uses correct terms.[3]

The language generation of the VAVETaM system is an application-oriented approach developed with the purpose to investigate a human-computer interaction system. The conception of the overall VAVETaM system and, in particular, the generation component are not intended to model a human-like behavior in terms of the underlying processes. Still, the task of the generation component is to generate understandable natural language. This is a task usually solved by humans. Therefore, insights from psycholinguistics and cognitive science (e.g., Habel, 1986, 1987; Levelt, 1989; Habel & Tappe, 1999) and natural-language generation systems that are cognitively oriented (especially, InC; Guhe et al., 2004; Guhe, 2007) are used in order to develop the generation component. For example, the components that Levelt (1989) identifies in a speaker—(1) conceptualizer, (2) formulator, (3) articulator, and (4) discourse model and situational knowledge—are reflected by the components of the VAVETaM system.[4]

[3]I chose this example for the purpose of illustration. I did not assess whether confusing of 'left' and 'right' happens frequently in assisting utterances produced by humans.

[4]However, a monitoring component, which is a part of the human language faculty (Levelt, 1989), was

Regardless of the approach towards natural language generation, different tasks in natural language generation require different kinds of knowledge. For example, solving the conceptualization task needs knowledge about how to organize the output and a knowledge source that contains information so that it is appropriate for verbalization.

Most existing generation systems are goal-driven; that is, they solve a task according to a communicative goal which is input to the system and usually called a 'communicative intention' (Hovy, 1993; Reiter, 1994; Carstensen et al., 2009). Goal-driven systems are usually used to generate longer texts. In goal-driven systems, the parts of the output must be put in an appropriate order (a task usually called 'linearization'; Levelt, 1982) to produce coherent natural language. To be able to do so, for example, rhetorical relations between output parts are specified. An influential theory for such a specification is the rhetorical structure theory (RST) (Mann, 1984; Mann & Thompson, 1988). Basically, the underlying idea of this theory is that coherence in a text is established by a number of relations that can hold between parts of a text. For instance, parts of a text can detail other parts of the text—a rhetorical relation called 'elaboration'. Rhetorical structure theory has been formalized by Hovy (1993). Another approach towards specifying parts of natural-language output is the use of schemas (McKeown, 1985). Schemas can be described as regular patterns texts usually follow. For example, academic articles usually begin with an introduction followed by a body of text that discusses the main contribution of the text (this example is taken from Reiter & Dale, 2000).

In the generation component of the VAVETaM system, appropriate timing is more important than elaborate planning of the output. The approaches discussed are not appropriate for the generation of assisting utterances because the utterances are governed by the haptic exploration movements of the map user. The output does not follow an organization that can be suitably described by schema-based approaches and cannot suitably be described as related parts in terms of the rhetorical structure theory.

The generation component of the VAVETaM system reacts to a map user's exploration movements in a virtual environment. Such systems are situated in virtual environments (Garoufi & Koller, 2010). Situated systems "connect language to the world" (Roy & Reiter, 2005, p. 2), which is, in the case of VAVETaMs, the virtual tactile map. The situatedness of the assisting utterances entails an additional task: A system that works in a situated manner has to solve the *'When to say?'* task in addition to deciding what to say and how to say it. For example, deictic utterances such as 'This is Abbey Road' are only suitable in situations when the particular street called 'Abbey Road' is currently haptically explored. Otherwise, the deictic reference fails, which potentially leads to confusion of the map user. While systems that produce static text have to linearize the output in the generation process without timing demands, a situated system has to decide taking the changing (exploration) context into account. This changing context also produces demands with respect to the timeliness of the output. When the exploration situation changes, the output needs to be reconsidered so that the most relevant information can

not implemented. Implementing such a component would have included the necessity to implement a speech-comprehension system and was not necessary to solve the task. Furthermore, only a very sparse lexicon was implemented (if the implementation deserves the name at all).

be articulated and no outdated information is given to the map user. Timeliness of the output is an important aspect in the VAVETaM scenario, as only a timely interaction ensures functioning reference of deictic expressions.

In conclusion, the literature agrees that the distinction between conceptualization and formulation is a reasonable solution for the development of natural-language generation systems. The architecture of most systems reflects this distinction. Propositional representations of the content of the utterances to be produced are suggested as interface between conceptualization and formulation. Most approaches towards discourse planning solve the task to generate longer texts in scenarios that are not situated. Therefore, these approaches are not suitable for utterance planning in the generation component of the situated VAVETaM system. A situated system has to solve the additional task to decide when information should be uttered. The situated systems discussed mostly focus on other tasks and/or scenarios than the VAVETaM system and, thus, are only partly relevant. As exceptions, the SOCCER system (Herzog & Retz-Schmidt, 1990) and the InC system (Guhe et al., 2004) are relevant because in these systems representations of events control the conceptualization of natural language.

4.3 The Haptic Exploration of Virtual Tactile Maps

A situated generation system has to react to the map user's exploration movements in a timely manner. To enable this timely reaction, such a system has to be controlled by appropriate input. As discussed, events have successfully been used to control the generation of assisting utterances in systems that provide simultaneous descriptions. An event-based approach is also adopted for the generation component of the VAVETaM system. Important events in the VAVETaMs domain are exploration events; that is, segments of the haptic exploration process the user performs to acquire information from the virtual tactile maps.

To acquire information from virtual tactile maps, the user moves the pen-like handle of the haptic device, as discussed in Chapter 2. As mentioned, the handle is attached to something that can be described as a reverse robotic arm. Force enabling the perception of virtual tactile maps is generated based on the position of the joint between the handle and the arm. This point is called 'interface point'. The map user moves the interface point on the virtual tactile map by moving the handle with the hand holding it. In the following, I call the movements that the map user performs with the handle and, thus, with the interface point, 'hand movements' or 'exploration movements'.[5]

The VAVETaMs scenario is partly dynamic: The map user explores a static virtual tactile map with dynamic hand movements. The hand movements of the map user are not random, but structured. This structuring of the hand movements is governed by how

[5] Most, but not all hand movements are suitably described by the term 'exploration movements'. For example, a map user can lift the device up from the map surface to get to a corner of the map, which is a hand movement, but not an exploration movement because the users do not explore the virtual tactile map with this movement. Therefore, I use both terms. In most contexts, their meaning is interchangeable.

humans organize their spatial knowledge (see Chapter 2) and by how the static virtual tactile map they explore is organized.

For example, the exploration of a potential landmark, such as a building, is based on hand movements that are specific: Users explore the borders of the potential landmark and they explore the inner area. If such hand movements are performed, this segment of the exploration can be detected as a meaningful event in the exploration. It can be described as the exploration of the particular potential landmark. Likewise, map users perform repeated movements on tracks (they move forth and back on the track), which can be detected as a track exploration.

The user performs exploration movements to acquire knowledge about a specific map object, such as a track, or about a part of a map object, such as a dead end a track forms. The map entity focused upon is in the *haptic focus* of the exploration; it is the entity which is the current subject of the map user's exploration. The haptic focus is the respective dynamically determined region, which allows to detect the entity focussed upon. Exploration movements as exemplified structure the exploration process.

As discussed by Habel and Tappe (1999) for the case of drawing events, events can be conceptualized on different levels as they have an hierarchical structure. An event can have subordinate events. Such a conceptualization is also appropriate for the conceptualization of exploration movements in VAVETaMs. For instance, to explore a track, the map user first moves the interface point into the area on the map representing this track. Then, suitable exploration movements are performed in the respective are. When the exploration is finished, the map user leaves the track. Together, these subordinate exploration movements can be described as track exploration, which is a superordinate event.

Human observers construct such superordinate event conceptualizations while the exploration movement is ongoing. For example, an observer of an exploration of a virtual tactile map is able to detect that the map user is exploring a street before the respective exploration event is finished. A situated assistance-generation system needs to be equipped with the same ability; that is, superordinate events need to be detected while the user is still performing them.

For general haptic perception in real environments, Lederman and Klatzky (1987, 2009) describe a set of exploratory procedures (EPs), which are typical patterns of haptic explorations with the hands. Exploratory procedures indicate that a certain kind of knowledge about the object that is explored is desired. For example, a lateral motion of one hand is associated with the goal to gain information about the texture of an object. Exploratory procedures are described for haptic-perception categories such as identifying the shape of a three-dimensional object or identifying the temperature of an object.

In VAVETaMs, users perform *map exploratory procedures* (MEPs) to acquire knowledge from the virtual tactile map. MEPs are a specialized version of EPs. They are designed to serve the purpose of categorizing the map user's hand movements during map exploration. The categories are inspired by videos of visualizations of explorations of virtual tactile maps.[6] The five MEPs are trackMEP, landmarkMEP, regionMEP, stop-

[6]As the prototype implementations discussed in Sections 4.7.5 and 4.8 show, these MEPs are suitable to control assistance generation.

MEP, and frameMEP. They are mainly distinguished by the sort of object explored with the procedure. For example, users perform a frameMEP to acquire knowledge about the frame of the virtual tactile map.

Map users do not only focus on complete map objects but also on parts of them (e.g., dead ends, which are parts of tracks). As the corpus study discussed in Section 3.2 shows, human assistants choose their assisting utterances with respect to whether objects or parts of objects are explored. For example, they inform the map user about properties of parts of map objects, instead of giving information about the complete object when they consider this more appropriate (with verbal utterances such as 'This is the dead end that Abbey Road forms to the left'). The conceptualization of the map user's exploration movements takes this into account by allowing such parts of map objects to be in the haptic focus.

In contrast to approaches that are based on triggering audio playback when map objects are touched, the event-based conceptualization of the map user's hand movements allows for a more reliable generation of suitable utterances. For example, if map users quickly cross a track with their hand movements, they are probably not primarily interested in this track. Consequently, providing verbal information associated to it would not be helpful in this situation, but rather confusing. In contrast, if map users explore a track with a trackMEP, giving verbal information about it and its local surroundings is certainly helpful (see Chapter 3).

4.4 The Conception of the VAVETaM System

As described, the natural-language generation component, which the research reported is focused on, is part of the VAVETaM system. A structure of this system is proposed in Habel et al. (2010) and Lohmann et al. (2010). As part of a larger system, the generation component interacts with other components. To describe the interaction with them, I discuss the overall structure of the VAVETaM system in this section. The proposed structure of the VAVETaM system is illustrated in Figure 4.1.

The *haptic device* presents virtual tactile maps (i.e., the device realizes a virtual presentation of models of maps stored on an attached computer system in a way that the map user can explore them). In the research presented in this thesis, the device used was a Sensable Phantom Omni. See Chapter 2 for a discussion of the haptic device and of virtual tactile maps and Section 4.3 for a discussion of how map users explore virtual tactile maps. The haptic device provides a stream of position data in the form of coordinates, which is input to the component that interprets the exploration movements of the map user.

The virtual tactile map that is presented with the haptic device is rendered on the basis of a central-knowledge component, the *Virtual-Environment Tactile Map* (VETM) knowledge base. The VETM knowledge base stores information on two layers: a *propositional layer* and a *spatial-geometric layer*. The spatial-geometric layer stores information about the spatial characteristics of objects based on coordinates. Even though the map is three-dimensional, the spatial-geometric layer of the VETM knowledge base is two-

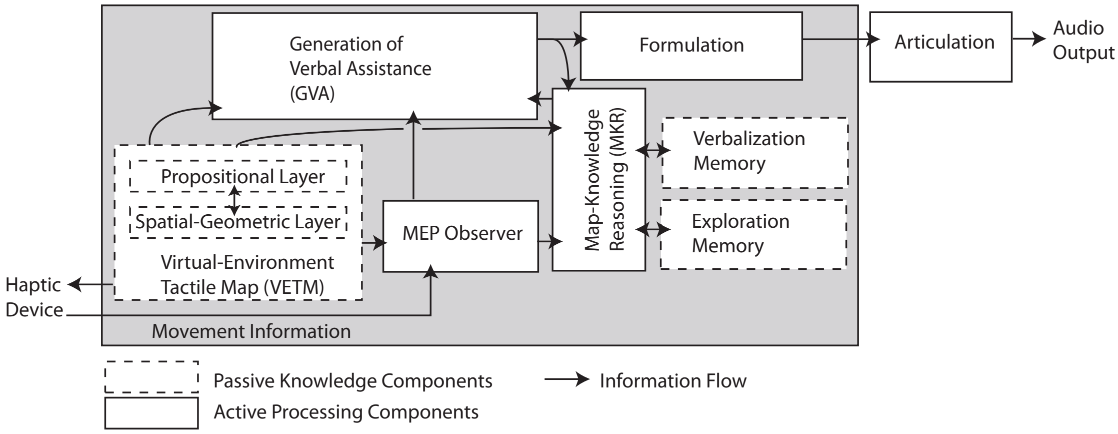


Figure 4.1: The proposed structure of the VAVETaM system (based on Lohmann et al., 2010).

dimensional. The propositional layer stores information needed for generating natural language. The representation formalism on which the propositional layer is based allows to establish links from propositional objects to spatial-geometric objects. For example, consider a map object called 'Hochstraße'. Information stored on the propositional layer allows to refer to this object as 'the street called Dorfstraße, which is left of Hochstraße'. This map object can be linked to a spatial-geometric object on the spatial-geometric layer of the VETM knowledge base described by the line from the coordinate pair A (say, (0, 100)) to the coordinate pair B (say, (100, 100)). The knowledge on the propositional layer is stored using the referential-nets formalism developed by Habel (1986). See Section 4.5 for a detailed discussion of the content of the propositional layer and an introduction to the formalism.

The *Map-Knowledge Reasoner* (MKR) component serves as memory component and keeps track of verbal and haptic information that the user has received. It controls two passive memory components, the *Verbalization Memory* and the *Exploration Memory*. As the names of the components suggest, the Verbalization Memory keeps track of the verbal information that the map user has received. The Exploration Memory keeps track of the the map user's exploration movements; that is, it keeps track of which parts of the map the user has explored. For the generation component, the Map-Knowledge Reasoner component is important to prevent unnecessary (and potentially disturbing) repetitions of utterances. Additionally, according to the conception of the VAVETaM system as discussed in Habel et al. (2010) and Lohmann et al. (2010), the Map-Knowledge Reasoner component has the task to anticipate which potential verbal information is most important and, hence, should be uttered before other information. Currently, the Map-Knowledge Reasoner component is not yet fully implemented. To test the prototype implementation of the generation component, a verbalization memory was created. This component is discussed in Section 4.7.5. It is sufficient to prevent unnecessary repetitions of utterances.

To be able to generate situated verbal assistance, the stream of position data from the haptic device has to be analyzed and conceptualized. The *Map-Exploratory-Procedures Observer* (MEP Observer) component is concerned with this task. The conceptualization of the exploration movements is feed into the generation component (discussed subsequently). The Map-Exploratory-Procedures Observer component categorizes the map user's movements to MEPs. Furthermore, if possible, the component specifies MEPs with the map object that is or the map entities that are currently in the haptic focus. In Section 4.6, MEPs and their specification are discussed in more detail.

The *Generation of Verbal Assistance* (GVA) component is the component concerned with two central tasks of situated natural language generation.[7] Specifically, it conceptualizes the assisting utterances taking the 'When to say?' task into account. The component receives a conceptualization of the user's hand movements from the MEP Observer component. The generation component selects the knowledge that is suitable for verbalization in an exploration situation from the propositional layer of the VETM knowledge base and prepares it so that it is appropriate for formulation. It interacts with the Map-Knowledge Reasoner component to prevent unnecessary repetitions of utterances. It sends preverbal messages to formulation. The inner workings of the generation component are discussed in more detail in Section 4.7. The *formulation* component generates natural language from preverbal messages. A prototype implementation of the formulation component is discussed in Section 4.7.5.

The *articulation* component produces speech. Currently, the interface between the formulation and the articulation component is text and we use standard synthesis software to generate verbal utterances from this text.[8]

4.5 Representing Knowledge for Verbalization

As discussed in the previous section, the VETM knowledge base specifying the structure and content of the virtual tactile map is a central component in the VAVETaM system and serves multiple purposes. On the one hand, it specifies the spatial-geometric layout to enable the presentation of virtual tactile maps. The virtual-environment character of the interface enables the user to experience the virtual tactile map as an analog representation of the presented area (for a discussion of analog and propositional representations see Chapter 2). The geometric specification is also needed to identify and specify MEPs. On the other hand, the VETM knowledge base supports the representation of map entities that can be named and described using natural language for generating assisting utterances.

The VETM knowledge base is a hybrid representation (see Chapter 2) that includes a spatial-geometric layer and a propositional layer. The propositional layer, which is in the focus of the following discussion, contains a propositional representation with a logical

[7] Earlier and later in the text, I mostly refer to this component as 'generation component' to increase readabilty.

[8] The system works with commercial voices from Acapela (http://www.acapela-group.com) and the open source text-to-speech (TTS) system Mary (http://mary.dfki.de and Schröder & Trouvain, 2003).

foundation. It represents semantic knowledge about map entities that can be used to generate assisting utterances in a manner appropriate for verbalization.

The representation that is used in the prototype of the generation component of the VAVETaM system and discussed in the following is proposed in Lohmann et al. (2011). Structuring the representation according to the formalism of *referential nets* (Habel, 1986) provides the necessary basis for linking to spatial-geometric information and accessing information corresponding to the current haptic focus. The following sections introduce the representation formalism.

4.5.1 Conceptual Sorts

The formalism that is used for the propositional layer of the VETM knowledge base is based on an order-sorted logic. Therefore, it allows to structure the domain of the representation according to the sorts of entities. The sorts are ordered by inclusion. The structuring of the map entities is important to select appropriate information for the generation of natural language. The identification of an appropriate set of *conceptual sorts* for the map domain is crucial for two reasons: (1) the sorts support the specification of predicates and relations for the generation of natural language and (2) they support the conceptualization of the hand movements the map user performs.

As discussed in Chapter 1, VAVETaMs are intended to be used for the acquisition of survey knowledge of urban areas. The conceptual sorts needed for the internal representation of the content of an urban map categorize entities that define a spatial setup with structures (such as streets) that enable locomotion. The conceptual sorts represent basic knowledge structures that are universal to representations of urban areas. Empirical research has shown that humans' internal representations of urban areas are structured in comparable categories. In Chapter 2, it was discussed how these categories map to the sorts proposed here (Section 2.4.3).

Note that while the conceptual sorts are generally suitable for urban-area maps used to acquire survey knowledge, they are not suitable for other domains of map use: A map used to communicate information about the size, position, and shape of the states of Europe requires a different set of conceptual sorts. In this case, for example, borders can be more relevant than tracks.

Figure 4.2 visualizes the part of the hierarchical structure of the sorts discussed in the following. The most general sort relevant for representing urban map content is called `map_entity`. A *map entity* is an object in the map that can be explored haptically or that can be specified by verbal assistances. Among the map entities, there are two main subgroups: Representational map entities (`rep_map_entity`), which represent objects present in the relevant fragment of the real world, and non-representational map entities (`nrep_map_entity`), which do not represent an object that is present in the relevant fragment of the world. For example, the frame of the map is such a non-representational map entity (in contrast to other map entities such as landmarks, the frame of the map is not a representation for an object in the real world). Knowledge about whether a map entity is representational is needed, because it is important for the language generation process. For example, it is needed to distinguish which of the following utterances should

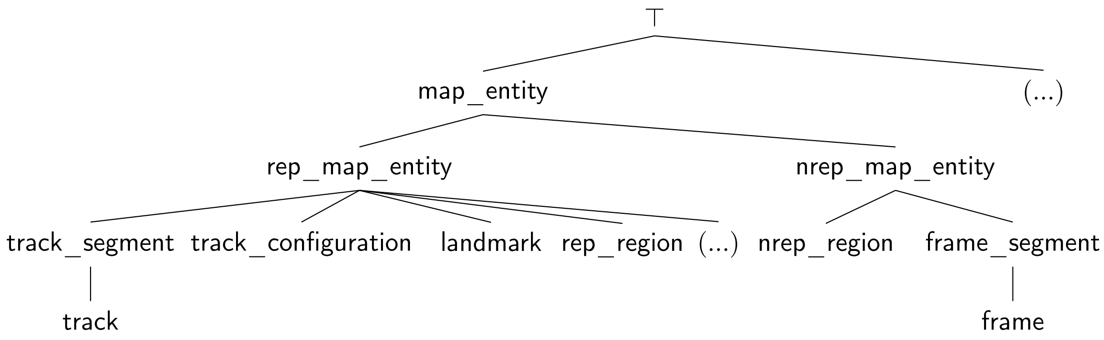


Figure 4.2: Part of the sort hierarchy for the urban map domain (based on Lohmann et al., 2011).

be generated: (1) 'This is Hochstraße. It is to the left restricted by the map frame' and (2) 'This is Hochstraße. It runs into Bergstraße to the left'. While (1) is appropriate in scenarios in which Hochstraße is not fully shown because the map is too small, (2) is appropriate in scenarios in which the real street ends in a T-intersection with another track.

Relevant sorts of representing map entities are those representing paths enabling locomotion (track, track_segment), such as streets, and those representing any kind of junction of such tracks (track_configuration). Usually, several track segments are part of a track. For example, dead ends, landmarks, and intersections can induce track segments between them and around them, which are explicitly represented if information for verbalization relating to these parts is included (a visualization of a map that includes track configurations is shown in Appendix D). Tracks are treated as a sub-sort of track segments as a simple track might consist of only one segment and, in this case, it is not justified to introduce two distinct objects into the representation. Furthermore, track segments might be complex, consisting of several smaller segments, just as tracks. Tracks are distinguished from other track segments, as they carry names. Therefore, they are represented as individual entities in the internal model that map users build up during their exploration. As they carry names, linguistic reference to tracks is rather simple. Track configurations can be intersections, T-intersections, corners in which streets meet, and so on.

Map entities representing potential landmarks such as buildings or trees (landmark), and map entities representing (named) areal objects such as squares or public parks (rep_region) require further sorts of representational map entities. As discussed in Chapter 2, humans structure their spatial mental models of urban areas on the basis of additional categories such as edges (linear objects that are not tracks, such as coast lines). On the maps that were investigated, no such objects are located. However, when maps are represented that contain such objects (to be described in the verbal assisting utterances), the introduction of according sorts is necessary.

Among the non-representational map entities are the frame and its parts (frame, frame_segment) and regions (nrep_region) defined relative to the frame of reference used in spatial descriptions (such as 'the upper part of the map'). In analogy to the track segments, the map frame is treated as a special case of frame segment. As the virtual tactile maps are bordered by a rectangular frame, the four sides are represented as four frame segments.

4.5.2 Referential Objects in Referential Nets

A referential net consists of a set of interrelated *referential objects* that represent entities of the domain of discourse (Habel, 1986). As the map entities serve as anchors for grounding the verbal assistance in the multi-modal setting described, the object-centered referential-nets formalism is particularly well-fitted as a representational framework (see Guhe, Habel, and Tschander (2003), for a discussion on how incremental conceptualization of natural language can be realized with this formalism). In Guhe et al. (2004) and Guhe (2007), the application or referential nets for natural language generation tasks is discussed. Figure 4.3 shows an example referential object. plm1 is an arbitrary identifier for the referential object representing a building with the name 'Rathaus'. As can be seen in the figure, edges connect the identifier of the referential object with expressions on both sides. The expressions on the left are attributes and the expressions on the right are designators.

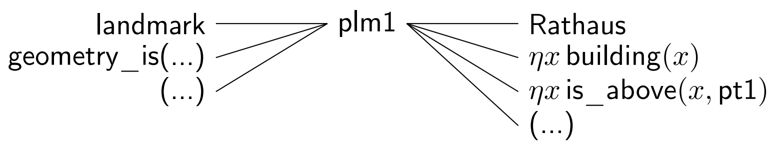


Figure 4.3: Referential object representing a house called 'Rathaus' (based on Lohmann et al., 2011).

Designators are terms of the underlying order-sorted logic. Complex designators are descriptions (e.g., ηx building(x) and ηx is_above(x, pt1)), whereas atomic designators are names (Rathaus). Descriptions can be constructed by using special description operators. Descriptions are of the form: *op var formula* with the description operators *op* $\in \{\iota, \eta, all\_t, some\_t\}$, the variables *var* $\in \{x, y, z, \ldots\}$, and *formula* a well-formed formula of predicate logic (Habel, 1982, 1986; Guhe, 2007).[9] The operator *op* indicates the cardinality of the referential object and the definiteness of the designation. The operators ι and *all_t* indicate that the description is definite. For example, ιx building(x) could be verbalized as 'the building' and *all_t* x building(x) could be verbalized as 'all buildings'. The operators η and *some_t* indicate that the description is indefinite. For example, ηx building(x) could be verbalized as 'a building' and *some_t* x building(x) could be verbalized as 'some buildings'. η and ι are used with single objects, *some_t* and *all_t* are used for classes of objects.

[9]See Russell (1905) and Linsky (2011) for a discussion of the ι-operator and definite descriptions.

In the prototype implementation of the generation component, only the description operators ι and η are used. In the example maps, there are no classes of objects because this leads to a variety of additional phenomena that need further investigation.[10] The referential-nets formalism has the capability to represent the semantics of utterances including references to classes of objects. Thus, it is possible to extend the generation component so that such utterances can be represented in the representational framework.

While designators represent knowledge allowing to refer to the respective objects in the world (Habel, 1986; Eschenbach, 1988), attributes of a referential object represent knowledge needed for inferences and reasoning. For example, the sort of a referential object is represented by an attribute assigned to it (such as landmark in Figure 4.3). The referential-net formalism allows for an assignment of additional attributes to a referential object. For the present purpose, two additional attributes are necessary to specify the mereological relation between different referential objects. The part_of attribute is used to express that a map entity is part of another map entity, as is the case for frame segments, which are parts of the complete frame. The parts attribute is the counterpart to the part_of attribute; it is used to represent that a map entity has one part or more parts. The latter is represented as list of parts—for instance, the frame has the frame segments as parts. Furthermore, the attribute geometry_is is used to link the propositional layer to a spatially defined object in the spatial-geometric layer.

See Habel (1986) for a formal specification of the referential-net approach and Guhe et al. (2004) for a more detailed discussion of the use of referential nets as representation formalism for natural-language generation systems.

The corpus study reported in Chapter 3 showed that human assistants do not distinguish between real-world objects and geometric objects. For example, they do not use 'the square' when they want to refer to a landmark with a square outline. Rather, they refer to the map objects in real-world terms such as 'the church'. The VETM knowledge base reflects this fact. Consequently, the real-world names of the objects on the map are represented exclusively.

4.5.3 Constant Symbols, Predicate Symbols, and Relation Symbols

The logical language used for building designators provides the usual signature of non-logical symbols. Due to the fact that the language is sorted, each symbol is assigned a specification regarding both arity and sorting.

Constant symbols represent names of objects. They are included to support the identification of objects corresponding to the message class 1.1 (see Chapter 3, for a discussion

[10] For example, it is an open question how classes of objects should be addressed efficiently in the assisting utterances regarding the verbalization history. It is plausible that when an assisting utterance such as 'This is Abbey Road. Above this street are *three buildings*.' was given in the discourse, and the map user starts exploring one of the three buildings, it is suitable to say 'This is the leftmost of the three buildings, it is the town hall.' In contrast, it is not suitable to construct this kind of utterance when the three buildings have not yet been introduced as complex in the discourse. In the latter case, an utterance such as 'This is the town hall. It is the leftmost of three buildings' would be much more appropriate.

of message classes). The signature underlying the logical language maps each constant symbol to the sort of object it might name. Typically, names are assigned to streets, buildings, parks, or squares. Constant symbols corresponding to such names are mapped to the sorts track, landmark, or rep_region.

Table 4.1: Some (nominal) predicate symbols for urban-area maps and the corresponding sorts (Lohmann et al., 2011).

Predicate Symbols	**Sort**
street	track
dead_end	track_segment
building, school_building, town_hall, church, tree	landmark
crossing, t_intersection, corner	track_configuration
region, square, park	rep_region
map_frame	frame
left_frame, top_frame, right_frame, bottom_frame	frame_segment

(Unary) predicate symbols represent properties of the represented entities. As the representing map entities inherit the properties of the real-word object they represent, several of the predicate symbols stand for properties of the underlying real-world object that are expressed by natural language nouns. Nominal predicates can support messages of the class identification 1.1 and are additionally needed to generate messages of the class 2.1 and 2.2, which specify the ends and junctions of tracks, and 3.2 specifying regions like squares and parks. Nominal predicates allow a finer grained specification of map entities. Thus, in the case of urban-area maps, predicate symbols such as building, school_building, town_hall, and church can be included and be mapped to the sort landmark (cf. Table 4.1).

Some of the predicate symbols listed in Table 4.1 relate to a geometric specification. For example, whether a track segment is a dead end can be determined based on the geometric specification of the track it is part of (if it is part of a track), the track segment itself, and interrelated track segments.

To facilitate the acquisition of survey knowledge, spatial information should be verbalized as discussed and evaluated in Chapter 3. Hence, additional predicate and relation symbols carry mainly spatial meaning (see Table 4.2).

The content of the propositional layer of the VETM knowledge base can be computed in a pre-processing step (the prototype implementation of the generation component follows this principle). This proposed computation is based on the geometric specification of the map objects and information about their interrelation. In the pre-processing, the propositional layer of the VETM knowledge base can also be controlled for consistency.

In Appendix C, an example referential net based on the symbol inventory discussed is shown. The net shown there represents the map visualized in Appendix D.

Table 4.2: Spatial predicate and relation symbols, sort frames, meaning, and supported message class.

Symbol	**Sort Frame**	**Meaning**	**Message Class**
is_above, is_left, (...), is_left_above, (...)	(map_entity, map_entity)	A map entity is above, is left of, (...), is left above, (...) another map entity.	1.2
is_between	(map_entity, map_entity, map_entity)	A map entity is between two map entities.	1.2
is_leftmost, is_topmost, (...)	(map_entity)	A map entity is the left-most, topmost, (...) map entity in the map.	1.4
has_left_limit, has_upper_limit, (...)	(track_segment, map_entity)	The left, upper, (...) end of a track segment is connected to another map entity.	2.1
is_left_part, is_upper_left_part, (...)	(track_segment, track_segment)	A track segment is the left, upper left, (...) part of another track segment.	2.1
is_in_track_config	(track_segment, track_configuration)	A track segment is part of a track configuration.	2.2
is_parallel_to	(track_segment, track_segment)	Two track segments are parallel.	2.3
is_located_in_nrepr	(map_entity, nrep_region)	A map entity is located in a non-representational map region.	3.1
is_located_in_repr	(map_entity, rep_region)	A map entity is located in a representational map region.	3.2

4.6 Conceptualizing the User's Exploration Movements

To be able to generate assisting utterances, the generation component needs to receive appropriate input representing the current exploration movements of the map user. The continuous stream of coordinates that the haptic device outputs has to be segmented to semantic (i.e., meaningful) units that are suitable to control assistance generation. According to the proposed structure of the VAVETaM system, this task is performed by the MEP Observer. This component outputs MEPs. The MEPs are, when adequate, specified with the map entities that are in the haptic focus. As dynamic input to the generation component, MEPs and their specification are important for the function of the generation component. In this section, I discuss MEPs and their specification.

4.6.1 Event-based Control of Utterances

For the generation of situated verbal assistance, the "symbolic realm" of words must be connected to the "non-symbolic realm of sensory-motor interaction" (Roy & Reiter, 2005, p. 3). A speaker observing and talking about a dynamic domain solves this task by "segmenting the stream of experience into meaningful units" (Habel & Tappe, 1999, p. 13). The speaker has to detect events in the stream of perceptions by appropriate segmentation of the observation. The control of assisting utterances in the VAVETaM system is based on such a segmentation. Meaningful segments in the exploration process (i.e., exploration events) are detected in the stream of coordinate-based hand movement data that the haptic device outputs.

In contrast, current approaches towards audio-tactile maps (see Section 2.3.3) are trigger-based. Speech playback is triggered when an object is touched. As mentioned in Section 4.3, trigger-based approaches have drawbacks in the situated and time-critical scenario: It is not possible to distinguish whether a map user wants to know something about a map entity or just touches it accidentally; for example, because it is located in the path users have to traverse to get to another map entity, which they want to gain knowledge about. Furthermore, trigger-based approaches do not have the capability to identify whether a map object, such as a track, or a part of a map object, such as a dead end, is focused upon.

As discussed in Section 4.3, exploration events that users' perform to acquire knowledge from a virtual tactile map have a temporal extend and an internal structure. Because exploration movements are conceptualized as events with a temporal extension, it is possible to control the generation of assisting utterances more versatile than in trigger-based approaches: Assisting utterances can be given as long as the exploration situation is ongoing and the utterance of outdated information can be prevented when the exploration situation has changed. In a trigger-based system, the information about the temporal extend of the respective exploration event is lost. Consequently, it is not possible to adapt the amount of information given for a particular map object and its surroundings to how long the map object is explored.

In the case of explorations of virtual tactile maps, a number of circumstances affect which information shall be given via natural language in an exploration situation: (1) how the map user explores the map (which MEP is performed), (2) which map entity is in the haptic focus, and (3) what has happened before. Thus, an *exploration situation* can be characterized by the MEP performed, the map entities that are in the haptic focus, and the exploration and verbalization history. With respect to the current hand movements of the map user, information as described by (1) and (2) is relevant. This information specifies the *current exploration situation*. MEPs have to be detected in the data that the haptic device outputs. Additionally, which particular map entity is in the haptic focus of the MEP has to be given to the generation component. Therefore, MEPs are specified with the map entities that are in the haptic focus, whenever this is appropriate.

Table 4.4: Types of MEPs and their specification.

MEP Type	**Required Specification**	**Optional Specification**
trackMEP	track-segment(s) identification or track-configuration identification	
landmarkMEP	landmark identification	
regionMEP		region identification
frameMEP	frame-segment identification	
stopMEP		

The track structure on the map is explored by (repeated) line following. By observing this characteristic motion, a trackMEP can be identified. As discussed, in addition to information about the exploration category, the generation component needs information about the objects in the haptic focus to select appropriate information. Correspondingly, a trackMEP is specified by the specific track segment(s) or track configuration that is in the haptic focus.[11]

In addition to track segments, potential landmarks are the subject of exploration. Their exploration is performed by a landmarkMEP. As for the trackMEP, to be appropriate input for the language generation component, the landmarkMEP is specified with the potential landmark in the haptic focus. An additional exploration category is the region exploration category. Some map regions might not be represented on the propositional layer of the VETM knowledge base that stores the information needed for verbalization (e.g., the upper left area of the map). Consequently, for the regionMEP, the specification of the region in the haptic focus is optional. A frequent and important exploration category is the exploration of the frame of the map. The corresponding exploration category is frameMEP. This category is specified with an identification of the frame segment touched. The stopMEP is an indicator that no exploration movements happen. No map entities are focused upon by a stopMEP and no specification is required.

As the user interacts at one specific point with the map, no more than one MEP can be performed at a time. As a result, the dynamic input to the generation component

[11]Note that, as discussed in Section 4.5.1, tracks are a sub-group of track segments.

are sequentially ordered MEPs and their specification. In the following examples, MEPs and their specification are exemplified and discussed in more detail. The first example discusses the conceptualization of an introductory exploration scenario. Furthermore, a notational convention for specified MEPs is introduced.

As discussed, MEPs are specified with the map entity or map entities in the haptic focus. Map entities can have parts, for instance, a dead end is a part of a street. Consequently, it is possible that more than one map entity gets in the haptic focus. Therefore, MEPs can be specified with multiple foci. One of these foci is the *primary focus* that indicates which map entity the map user primary explores. Example 1 illustrates the conceptualization of an example exploration with primary foci.

The map entities that are in the haptic focus, but not in the primary focus, are in the *secondary focus*. While only one map entity can be in the primary focus, multiple map entities can be in the secondary focus. Example 2 illustrates the conceptualization of an exploration situation with a secondary focus.

The conceptualization of exploration situations involving track configurations is discussed in Example 3. Finally, Example 4 shows a case in which a track_segment that is part of a track is in the primary haptic focus. To exemplify the use of MEPs and their specification in a real map-exploration scenario, I discuss a conceptualization of a part of a map exploration a participant of the experiments discussed in Chapter 3 performed.

Example 1: An Introductory Example (Primary Haptic Focus)

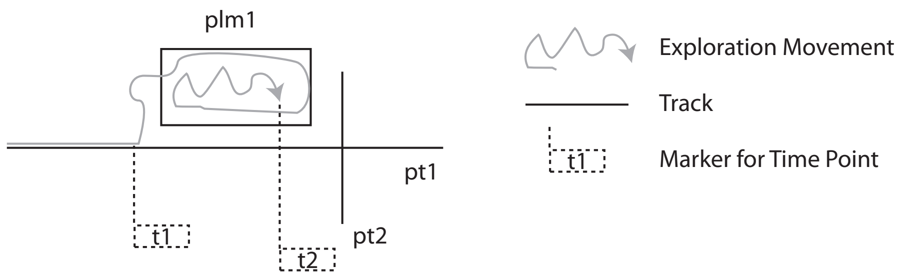


Figure 4.4: Example exploration 1 with primary haptic foci.

Figure 4.4 shows a visualization of an example exploration of a tactile map.[12] pt1, pt2, and plm1 are identifiers for the referential objects on the propositional layer of the VETM knowledge base. The grey arrow represents the hand movement of the map explorer. Two distinct time points are marked with dashed lines and the labels t1 and t2. The figure shows an exploration of a part of the track pt1 until shortly after time point t1. Subsequently, the explorer makes a hand movement to the landmark plm1 and starts to explore this landmark. From the beginning (represented on the very left) till leaving

[12] The track segments that surround the intersection and the intersection itself are omitted from the figure for the purpose of a clear presentation. Refer to Example 3, for a discussion.

the track, the track pt1 is followed, which is a characteristic movement to explore tracks. Therefore, in this phase of the exploration, the hand movement can be characterized by a trackMEP. A human observer as well as an artificial system detecting MEPs needs some time for a proper on-line classification of the map user's hand movement. Therefore, after an initial phase needed to detect the trackMEP after the track exploration is started, the input to the generation component is trackMEP. In the short phase between the time point at which the map explorer leaves pt1 and before the movements clearly indicate an exploration of the landmark plm1, no MEP is detectable. Then, at a certain point, the hand movement can be classified as landmarkMEP.

As discussed earlier in this section, MEPs have to be specified with the identifier of the map entities in the haptic focus. In the example discussed, this is one map entity for each MEP (Example 2 and 3 illustrate and discuss example situations with multiple foci). This MEP is specified with the identifier of the track in the haptic focus, which is pt1. In the following, as a notational convention I add the specification of the MEP in parentheses after the MEP, for example, trackMEP(pt1). Shortly after time point t1, the map user stops exploring the track, and initially no MEP is detectable. Therefore, the generation component does not receive any input. It takes some time to detect the landmarkMEP after the map user entered the landmark plm1. After the MEP is detected, the input to the generation component changes to landmarkMEP(plm1). Table 4.5 provides an overview about the input to the generation component at the time points t1 and t2.

Table 4.5: The input to the generation component at time points t1 and t2 resulting of a conceptualization of the movements discussed in Example 1.

Time Point	**Input to the GVA**
t1	trackMEP(pt1)
t2	landmarkMEP(plm1)

It is possible to visualize the stream of MEPs on a timeline, as well. Figure 4.5 shows such a visualization for the example discussed in this section.

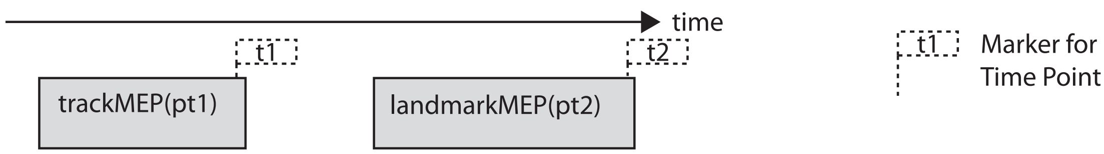


Figure 4.5: The input to the generation component represented on a timeline.

By receiving the conceptualization of the map user's hand movements in the form of MEPs and their specifications, the generation component is able to generate helpful assisting utterances. (1) and (2) are examples for such utterances, assuming that the track pt1 is called '3rd Avenue', the track pt2 '42nd Street' and the landmark plm1 is called 'town hall'.

(1) This is 3rd Avenue. Above this street, there is the town hall. (...)

(2) This is the town hall, which is left above the crossing of of 3rd Avenue and 42nd Street. (...)

Example 2: Situations with a Secondary Haptic Focus or Multiple Secondary Haptic Foci

The VETM knowledge base allows for spatially overlapping map entities. Hence, in some exploration situations, the specification of an MEP with a single haptic focus is not sufficient to conceptualize the exploration situation adequately. This and the following examples will mostly be concerned with explorations of the track network. Due to the fact that track segments can be parts of tracks or other track segments, the track network provides comprehensible examples for the input specification for the generation component. For example, several track segments can be part of a track. As discussed in Section 4.5, dead ends, landmarks, the ends of tracks and intersections can induce track segments between them and around them, which are explicitly represented if information for verbalization relating to these parts is available on the propositional layer of the VETM knowledge base.

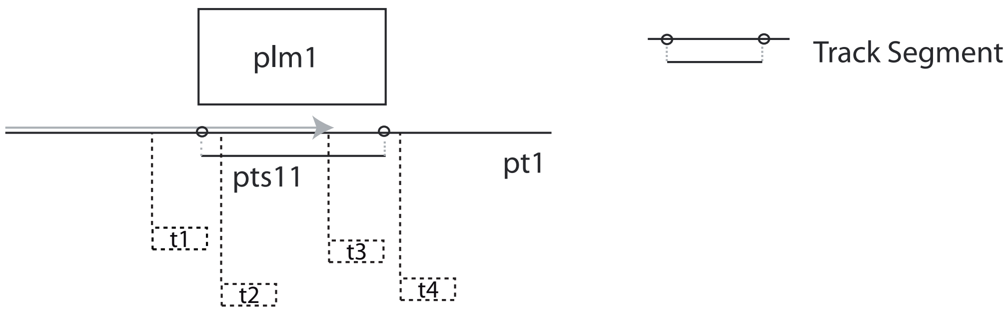


Figure 4.6: Example exploration 2, involving multiple haptic foci.

Figure 4.6 shows an exploration situation in which a track segment is part of a track and partly overlaps the track. The part below the potential landmark is represented as a unique track segment with which information for verbalization is associated. This information is, for example, used to generate utterances such as (3) (assuming that the potential landmark plm1 is called 'town hall', again).

(3) Now, you are below town hall.

The user explores the virtual tactile map with hand movements along the track pt1. At the time point t1, the input to the generation component is, in compliance with the discussion above, trackMEP(pt1). The map user explores pt1 further and at time point t2, he or she enters the track segment pts11[13], which is located below the potential landmark plm1. At this time point (or a little later, depending on whether there is time needed to recognize that the user has entered track segment pts11, see Section 4.7.5, for a discussion), the input to the generation component changes to trackMEP(pt1, [pts11]).

I use the following convention to notate primary and secondary foci: The primary focus is specified with its identifier after the opening parenthesis, and the secondary focus or the secondary foci are specified with their identifiers in squared brackets. If more than one secondary focus is specified, this can be done as follows: trackMEP(pt2, [pts21, ...]) (the dots stand for one or more additional secondary foci). The order of the list of identifiers specifying secondary foci is arbitrary and not used in further computation.

Back to the example, the representation does not change until the user of the map leaves track segment pts11 again. Then, if the user of the map stays on track pt1, the input to the generation component changes back to trackMEP(pt1) (circa at time point t4 in Figure 4.6). Table 4.6 shows an overview of the MEPs and specification at the time points marked in Figure 4.6.

Table 4.6: The input to the generation component for Example 2.

Time Point	**Input to the GVA**
t1	trackMEP(pt1)
t2 and t3	trackMEP(pt1, [pts11])
t4	trackMEP(pt1)

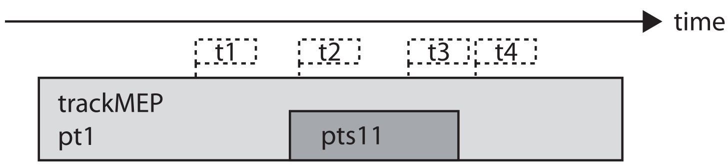


Figure 4.7: Input to the generation component with a secondary focus represented on a timeline.

Figure 4.7 shows a visualization of the input to the generation component. The large grey box in a lighter grey shows the MEP, which does not change during the example. The identifier of the map entity with which the MEP is specified is shown in the bottom left part of the figure (pt1). The small box colored in a darker grey represents the secondary focus (with the identifier pts11), which is added to the MEP somewhen before or at t2

[13]The string pts is used with an index as identifier for track segments on the propositional layer of the VETM knowledge base.

and still current at t3. Between t3 and t4, the secondary focus becomes outdated and the specification of the MEP is changed back to trackMEP(pt1).

Example 3: Situations Which Involve Track Configurations

As discussed in Section 4.5, track configurations are constellations of at least two tracks. Track configurations can, for instance, be crossings or T-intersections. On the propositional layer of the VETM knowledge base, track configurations are represented as objects that consist of track segments. Therefore, exploration situations that involve track configurations do not impose special requirements on the classification and specification of the map user's hand movements.[14]

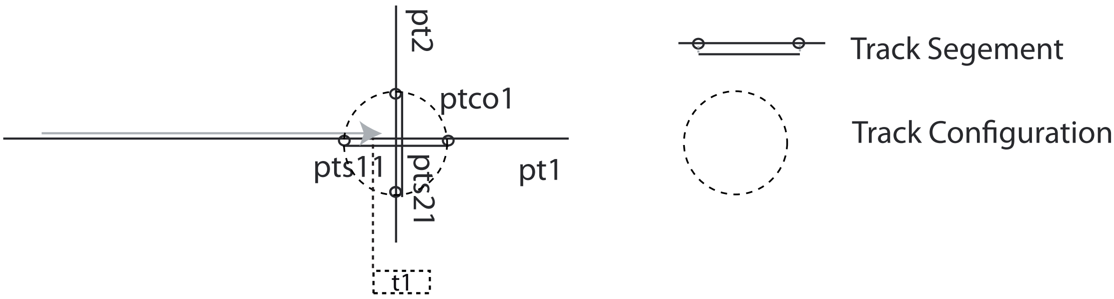


Figure 4.8: Exploration situation of Example 3, involving a track configuration.

Figure 4.8 shows an exploration situation in which a crossing of the tracks pt1 and pt2 is involved. The track configuration is marked by the dashed circle and has the identifier ptco1.[15] The track configuration consists of two track segments: the track segment pts11 as part of the track pt1 and the track segment pts21 as part of the track pt2. The principles of focus detection do not differ from those discussed in Example 2 for other track segments in the secondary haptic focus. The conceptualization of the map user's hand movements at time point t1 is therefore either trackMEP(pt1, [pts11]) or trackMEP(pt1, [ptco1]). In the former variant, the generation component detects internally that pts11 is part of the track configuration ptco1 and replaces the identifier in the specification of the MEP with the identifier of the track configuration the track segment is part of. This is enabled by the VETM knowledge base, which explicitly represents track segments that are part of track configurations using the parts attribute (see Section 4.5.2, for a discussion, and Appendix C, for examples). This replacement

[14] A track configuration has a least two corresponding track segments, which are part of the tracks that constitute the track configuration. As Figure 4.8 shows for the case of a crossing of two tracks, no distinct point of intersection is explicitly represented on the propositional layer. A major advantage of designing the representation like this is that no special movement-conceptualization mechanisms have to be developed for explorations of track intersections: the general mechanisms for track segments are sufficient to cover explorations of track configurations.

[15] The string ptco is used with an index as identifier for track segments on the propositional layer of the VETM knowledge base.

results in an identical behavior of the language generation component for both input variants. The generation component processes both variants in the same manner. Given one of the inputs to the generation component, utterances such as (4) (assuming the tracks have the names '42nd Street' and '3rd Avenue') are generated.

(4) Here, 42nd Street and 3rd Avenue intersect. (...)

Example 4: Situations in Which Parts of Tracks are in the Primary Haptic Focus

If the track network is explored, in most cases, tracks are the map entities that are in the primary haptic focus. However, in some exploration situations, a part of a map entity is in the primary haptic focus.

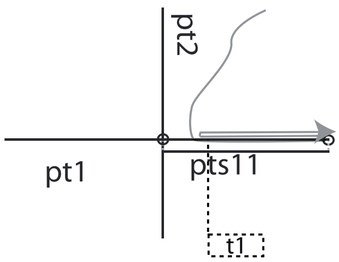


Figure 4.9: Example 4, involving a track segment in the primary focus (the visualization of the track segments induced by the track configuration and the track configuration itself are omitted to make the visualization clearer).

Figure 4.9 shows an example. The map user moves on the track from a region above the track and repeatedly moves along the right part of the track pt1, which forms a dead end. The dead-end part of track pt1 is represented with the identifier pts11. In situations like this, the exploration movements of the user allow the assumption that he or she is primarily interested in a part of the map entity. Therefore, if the part of the track is explicitly represented in the VETM knowledge base (pts11 in this example), it is suitable to verbalize information associated with this map entity. Utterance (5) is an example for a possible utterance, again assuming that pt1 is called '42nd Street'.

(5) This is the dead end that 42nd Street forms to the right. (...)

In the case of the example presented, the map user focuses on the track segment pts11. However, it is still suitable to talk about the track pt1 as well. Consquently, the track should also be part of the specification of the MEP. Therefore, at the time point t1, the conceptualization of the map user's movements should be trackMEP(pts11, [pt1]).

4.6.2 A Manually Annotated Conceptualization

As reported in Chapter 3, three experiments were conducted with the purpose to evaluate VAVETaMs. In these experiments, blindfolded sighted and visually impaired participants explored virtual tactile maps and received assisting utterances that were controlled by the experimenter. In some of the experiments, the map-exploration movements of the participants were recorded using screen-recording software. On the basis of one of these video recordings, conceptualizations of the exploratory hand movements of some of the map user's were created as sample input for the generation component. This was done by manually annotating the MEPs that were detectable in the screen recordings.[16] A part of one of these example conceptualizations is discussed in this section to illustrate how longer parts of map explorations can be segmented with specified MEPs.

Table 4.7: Manually created conceptualization of a map exploration: MEPs and specification. Refer to Figure 4.10, for a visualization of the map explored by the participant including the identifiers used in the conceptualization.

Time in Seconds	**Input to the GVA**
6–8	trackMEP(pt1)
8–16	trackMEP(pt2)
16–18	trackMEP(pt1)
18–27	trackMEP(pt2)
17–28	trackMEP(pt2, [ptco4])
28–29	trackMEP(pt3)
29–34	trackMEP(pt3, [pts33])
34–38	trackMEP(pt3)
38–39	trackMEP(pt3, [ptco4])
39–41	trackMEP(pt5)
41–44	trackMEP(pt5, [ptco2])
44–45	trackMEP(pt4, [ptco2])
45–52	trackMEP(pt4)
52	trackMEP(pt4, [pts44])
52–53	trackMEP(pt4)
54	trackMEP(pt4, [ptco2])
55–61	trackMEP(pts55, [pt5])
61	trackMEP(pt5)
61	trackMEP(pt5, [ptco1])
62–66	trackMEP(pt1)
66–67	trackMEP(pt1, [pts12])
67–71	trackMEP(pt1)
71–72	trackMEP(pt1, [pts11])

Table 4.7 presents a manually created conceptualization of the movements of the first 72 seconds of an experiment with a visually-impaired participant of Experiment 2. The

[16]The ANVIL software was used to annotate the videos (Kipp, 2001).

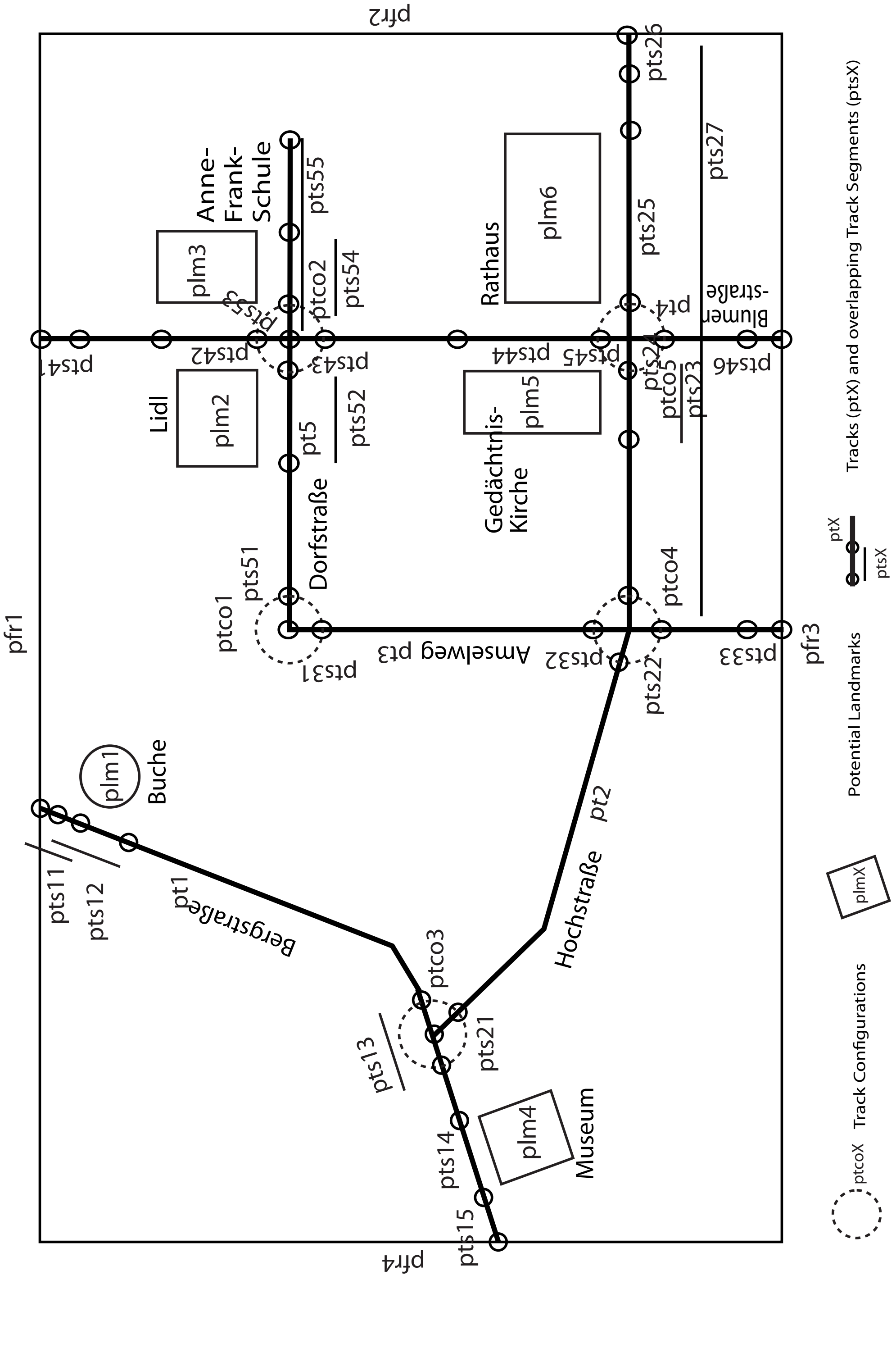


Figure 4.10: Example map used in the experiments and in the prototype implementation (this figure is identical to that shown in Appendix D and repeated here for convenience).

conceptualization was created by the author. The left column of the table shows the time period of the MEP in seconds. The right column shows the MEP identified and its specification using the notation introduced above. Figure 4.10 shows a visualization of the map used in the experiment and the identifiers used for the specification of the MEPs in Table 4.7 and in Appendix C, the corresponding referential net is presented. As can be seen, during the time period shown, only the track structure is explored by the participant. This manual annotation was used for initial testing of the prototype of the generation component. With this input, the prototype of the generation component generates assisting utterances comparable to those tested in the experiment. In Section 4.7.5, a more systematic approach of testing the generation component is described.

4.6.3 Summary: Conceptualizations of Map Explorations

In this section, map exploratory procedures (MEPs) were discussed as a basis for a conceptualization of a map user's hand movements. The types of MEPs that were suggested were trackMEP, landmarkMEP, regionMEP, frameMEP, and stopMEP. Only one MEP can be performed at a time leading to a sequential stream of MEPs while the user is moving on the map. However, it is possible that during time periods no MEP is detected—either because the detection fails or because the user is not performing a movement that indicates that he or she wants to gain information about a specific map entity. To fully capture the exploration situation in a manner sufficient to generate assisting utterances, MEPs can be specified with map entities that are in the haptic focus. For some types of MEPs, such as trackMEP and landmarkMEP, the specification is required in order to conceptualize the map user's current exploration situation. For others, such as regionMEP, the specification is optional. It is possible that more than one map entity is in the focus of haptic exploration, as different map entities can spatially overlap. For example, a track segment (for example a dead end) can be a spatially overlapping part of a track. In most situations, the exploration movements indicate which map entity the user primarily explores. To capture this information, MEPs can be specified with two kinds of haptic foci: Exactly one map entity can be in the primary haptic focus, other map entities whose associated information is currently suitable for verbalization are in the list of secondary haptic foci. The prototype implementation of the generation component was tested with a conceptualization of the hand movements like discussed as input (described in more detail in Section 4.7.5). With this input, the implementation of the generation component generates assisting utterances comparable to those discussed in Chapter 3.

4.7 The VAVETaM Generation Component (GVA)

4.7.1 Subcomponents and Their Interaction

The generation component solves the task of determining appropriate content for utterances in an exploration situation. It solves the task to decide when to say what. According to the proposed structure of the VAVETaM system (see Section 4.4), the generation component interacts with various other components: (1) it receives the concep-

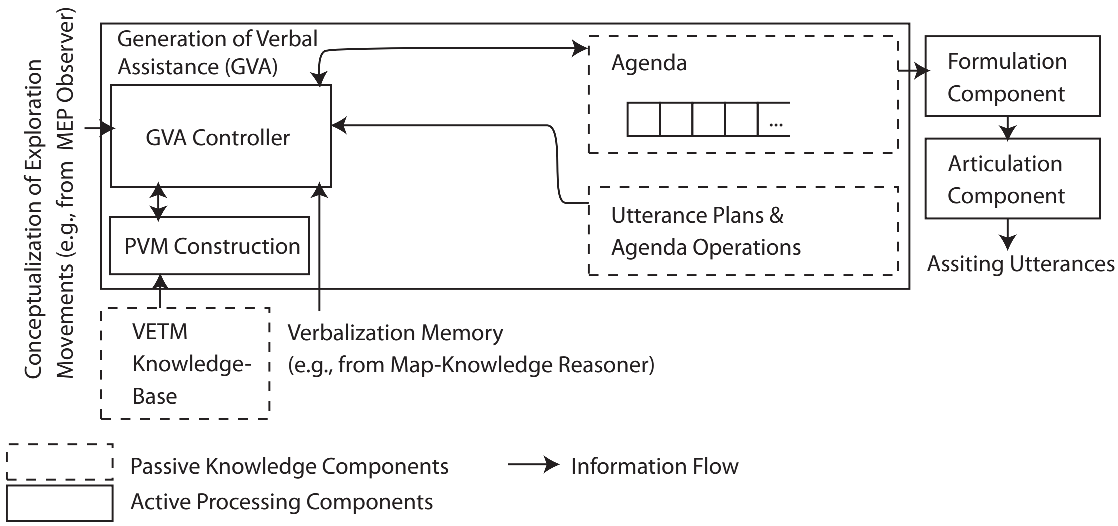


Figure 4.11: The structure of the generation component.

tualization of the user's movements (MEPs and specifications) from the MEP Observer; (2) it accesses the propositional layer of the VETM knowledge base in order to retrieve information that is suitable for verbalization; (3) it interacts with the Map-Knowledge Reasoner component, which keeps track of the exploration and verbalization history; and (4) it sends semantic representations in the form of preverbal messages (PVMs) to the formulation component.

The generation component consists of several subcomponents which are visualized in Figure 4.11. The *GVA Controller* controls the execution of other processes by controlling the *Agenda*, which is an ordered list of preverbal-message representations of utterances. Once the generation component receives a (specified) MEP that represents the current exploration situation from the Map-Exploratory-Procedures Observer, it looks up *Utterance Plans & Agenda Operations*. Utterance Plans & Agenda Operations specify which information is suitable to express in a given current exploration situation and where it should be placed on the Agenda. Where preverbal messages are placed on the Agenda is important with respect to appropriate timing of the messages (and thus, important with respect to the "When to say?" task in situated natural language generation). The first preverbal message on the Agenda is passed on to formulation as soon as articulation is ready to articulate a new preverbal message. The *PVM Construction* component searches an utterance plan that allows to construct preverbal messages that are appropriate in the exploration situation.

The following steps are performed until an utterance is articulated. First, the conceptualization task is solved resulting in a preverbal message of the utterance. Then, the utterance is formulated and articulated.[17] Currently, the architecture does not allow for

[17]In natural language generation, the problem that the result of conceptualization is possibly not re-

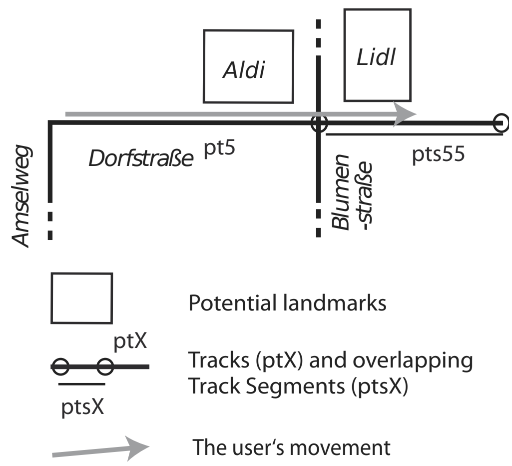


Figure 4.12: Visualization of a Part of a Virtual Tactile Map.

a feedback of information from the formulation or articulation components solving the formulation task to the component solving the conceptualization task.

In the structure of the VAVETaM system one important task of the Map-Knowledge Reasoner component is to store information needed to prevent unnecessary repetitions of utterances. In the prototype implementation of the generation component, a basic verbalization memory fulfills this task. For example, if the map user has previously explored the track pt5 and already received the information that the buildings 'Lidl' and 'Aldi' (see Figure 4.12) are above the track, adding a preverbal message containing this information to the Agenda is prevented and the user is given other information (or none, if no suitable information is available). The implementation of the verbalization memory was necessary to test the component.

4.7.2 Formulation and Articulation

("That is" | "This is" | "Here is" | "Now you're at")
(again)?
(Identification: Name or definite referring expression)

Figure 4.13: Literal translation of the template for a German identification message.

In the generation prototype, formulation of preverbal messages is realized with a template-based approach (see Section 4.2, for a discussion). The formulation component uses a set of sentence templates, which consist of partial formulations and gaps to

alizable with the linguistic resources available is called 'generation gap' (Meteer, 1991). Due to the restricted domain, it is ensured that all preverbal messages that are constructed can actually be formulated.

fill with appropriate information. Additionally, a lexicon stores knowledge about natural language expressions that can be used to inform the map user about spatial situations. Figure 4.13 shows a simple template used for the generation of identification messages.[18] Of the four utterance parts shown in the first line, one is chosen randomly enabling some variation in the utterances. If the memory component has marked the preverbal message as a repetition of a previously articulated utterance, a marker word is placed in the sentence (here: 'again'). Then, the sentence is completed by either uttering the name of the object in focus or by uttering a referring expression, as specified in the preverbal message.

The resulting text is synthesized in the articulation component using text-to-speech (TTS) software. Currently, the generation system works with different synthesis components. We tested commercial voices from Acapela[19] and the open source text-to-speech (TTS) system Mary[20].

4.7.3 Agenda

The Agenda is an ordered list of preverbal messages. The first preverbal message on the Agenda is formulated and articulated when the formulation component is idle; that is, when currently no utterance is articulated.[21] There are the following operations on the Agenda: (1) appending a preverbal message to the Agenda, (2) putting a preverbal message in front of the Agenda, (3) removing a preverbal message from the Agenda, and (4) clearing the Agenda. The Agenda enables to construct a plan for the next assisting utterances while the exploration is ongoing. If the exploration situation changes, only the planned utterances that are affected of this change (e.g., because they are outdated) are revised. Consequently, the Agenda allows for a incremental determination of the next utterances (incrementality is, e.g., discussed in Guhe, 2007).

4.7.4 Utterance Plans & Agenda Operations

An important part of conceptualization is to determine the content of utterances and to determine an order for expression. For an overview of classical approaches towards discourse planning see Section 4.2, Reiter and Dale (2000), and Lemon (2011). As discussed, these approaches are not applicable in the VAVETaM system. For assistance generation in virtual tactile maps, it is most important to express time-critical information firstly and to be quickly able to adapt to the current exploration situation. Classical approaches towards discourse planning are concerned with longer (text) output with stronger relations between the parts of the (text) output and are not focussed on time-critical situated generation. They are, thus, not suitable for the generation component of the VAVETaM system.

[18]Note that the system is implemented in German and the word order is a word-by-word translation; the ordering of elements leads to grammatically correct German sentences.

[19]http://www.acapela-group.com

[20]http://mary.dfki.de and Schröder and Trouvain (2003).

[21]In the prototype implementation, an additional delay was specified after a preverbal message was uttered because this let to more natural articulation.

For VAVETaMs, timely interaction of assisting utterances and the user's exploration movements is highly important. In contrast to other domains of natural language generation, achieving coherency of the output is unproblematic due to the tight coupling of exploration situation and the assisting utterances. This tight coupling ensures coherency of the output (e.g., that there are no topic shifts). It is important that the output adapts to changes of the current exploration situation (e.g., when the map user's haptic focus changes to another map entity) as quickly as possible. Which verbal information is helpful depends on the the sort of the map entity in the primary haptic focus. The type of the MEP indicates which sort the explored map entity has, for example, potential landmarks are explored with a landmarkMEP. Therefore, utterance plans are associated to MEPs and changes in the secondary haptic focus. In addition, knowledge about where the assisting utterances are put on the Agenda is associated with utterance plans. This is important to ensure that time-critical utterances (for instance, deictic utterances such as 'This is ...') are produced firstly.

Changes in the current exploration situation (indicated by changes of either the current MEP or its specification) are associated with utterance plans. Utterance plans are stored as lists of *potential messages* and construction rules. *Construction rules* specify the selection and combination of information for verbalization from the VETM knowledge base. For example, the construction rule for an utterance informing the user about geometric relations of two tracks (message class 2.3, see Chapter 3) includes the following information: Of all descriptions that contain the is_parallel_to symbol and are associated to the object that is currently in the haptic focus, one is selected for each relation with another object. For example, if track pt3 called 'Blumenstraße' is parallel to track pt4 called 'Amselweg', two descriptions which represent parallelism of these two tracks are associated to track pt3: (1) ηx is_parallel_to(x, pt4), which corresponds to utterances such as 'Amselweg is parallel to Blumenstraße', and (2) ηx is_parallel_to(pt4, x), which corresponds to utterances such as 'Blumenstraße is parallel to Amselweg' (see Appendix C, for the referential net to which this examples relate). Only one of these descriptions is selected. Currently, which one should be selected is not systematically investigated and selection depends on the specification of the utterance plan.

In the following, I discuss some example utterance plans for exploration situations in which a new MEP starts. With a trackMEP, it is associated that firstly, an utterance of the identification message class should be constructed (message class 1.1). Such a message is constructed by either stating the name associated to the map entity in the haptic focus (e.g.,'Dorfstraße') or by choosing or constructing a referring expression that allows for definite identification (e.g., 'The dead end which Dorfstraße forms to the right').

After the map entity in the haptic focus is identified, if available, information about geometric relations such as parallelism with other linear objects on the map (message class 2.3) is selected from the VETM knowledge base, followed by information about spatial relations (message class 1.2) with other map entities. Subsequently, the construction of a preverbal message that informs the user about the extension of the track segment (message class 2.1) in the haptic focus is tried, followed by information about crossings the track segment has (message class 2.2).[22] It is tested whether the VETM knowledge base contains suitable information for each of these construction rules. If it does, a preverbal message is constructed by the PVM Construction component and added to the Agenda unless the memory component prevents the preverbal message because it has been generated shortly before.

If an intersection is in the primary haptic focus, it can be identified by stating the names of the streets that form the intersection, with an utterance such as: 'This is the intersection between Blumenstraße and Hochstraße'. In order to construct this utterance, PVM Construction selects the two designations $\eta x\, \mathsf{is\_in\_track\_config}(\mathsf{pt2}, x)$ and $\eta x\, \mathsf{is\_in\_track\_config}(\mathsf{pt4}, x)$ and combines them to an utterance (which can be represented as $\iota x\, [\mathsf{is\_in\_track\_config}(\mathsf{pt4}, x) \wedge \mathsf{is\_in\_track\_config}(\mathsf{pt2}, x) \wedge \mathsf{crossing}(x)]$).

A change of the exploration situation to a landmarkMEP is associated with the following list of potential messages (message classes given in brackets): identification (1.1), spatial relation (1.2)[23], containment (3.2), and location on the map (3.1). Likewise, potential messages are associated to the changes of secondary foci during an MEP, for instance, when a track segment enters the secondary focus.

The construction of the potential messages according to the utterance plans is tried and, if successful, the resulting preverbal messages are put on the Agenda (after consulting the memory component). Where they are put on the Agenda is specified by *Agenda Operations*. This is especially important in situations in which a secondary focus is added to a trackMEP, because the corresponding information is time critical and should be put in front of the Agenda. Currently, when an MEP changes, the Agenda is cleared. Then, the list of preverbal messages that could be constructed based on the utterance plans associated with the type of the current MEP are put on the Agenda. When a second focus is added to an MEP and the construction of according preverbal messages is possible, those messages are put in front of the Agenda because they are usually time critical. When a secondary focus becomes obsolete, corresponding preverbal messages are deleted from the Agenda. In Section 4.7.6, it is exemplified how utterances are generated in the prototype of the generation component on the basis of utterance plans as discussed.

4.7.5 A Prototype Implementation of the Generation Component

A prototype of the generation component was implemented in Java in order to test its technical appropriateness. The output of the generation component needs to fulfill the following constraints: (1) situatedness, (2) appropriate order, and (3) completeness. Constraint 1 means that the assisting utterances are given in a timely manner at appropriate time points. Most importantly, only information associated to map entities that

[22] The order in which messages are generated is not evaluated empirically. According to the conception of the VAVETaM system as discussed in Section 4.4, the Map-Knowledge Reasoner component is able to anticipate the importance of information for the map user. Therefore, this component—which is an open issue—would be able to bring the preverbal messages in an appropriate order according to the importance for the user.

[23] In the prototype implementation, this message class is modeled by two potential messages: first, the construction of a preverbal message about spatial relations with other tracks is tried and then the construction of a preverbal message about spatial relations with landmarks with landmarks is tried.

are currently in the haptic focus should be given. Additionally, unnecessary repetitions of utterances shall be prevented. According to constraint 2, preceding any other verbal information, map entities that are in the primary haptic focus need to be identified (with a message of the class 1.1). Constraint 3 states that the generation system should be able to construct all messages of the classes tested in the experiments in Chapter 3, given the availability of adequate information in the VETM knowledge base. Furthermore, the system has to be robust against different kind of inputs (e.g., if map users explore the map either very fast or very slowly or if the MEP Observer component reacts either fast or slowly).

The generation component is intended as a part of the VAVETaM system, as discussed in Section 4.4. The generation component interacts with other components, most importantly with the propositional layer of the VETM knowledge base, the MEP Observer and a memory component (the Map-Knowledge Reasoner, according to the conception of the VAVETaM system discussed in Section 4.4). The goal of the prototype implementation was to show that verbal assistance as proposed in Chapter 3 can be generated by an artificial system. The VETM knowledge base provides a representation of the map based on the referential-nets formalism, which contains information needed to construct assisting utterances (see Appendix C, for the representation of the map used in the following). Referential nets are input to the VETM knowledge base as XML-based representation. The XML representation allows easy editing of the propositional layer of the VETM knowledge base. Furthermore, it provides a flexible interface enabling to share knowledge with other components.

A memory component is needed to prevent unnecessary repetitions of utterances. A basic verbalization memory was implemented to solve this task. The version that was implemented allows to test the component. It does not keep track of the exploration and does not allow for reasoning about the map user's information needs.

To be able to produce situated and timely assisting utterances, the generation component receives (specified) MEPs from the MEP Observer component. As described in Section 4.4, the detection of haptic exploration events is not in the focus of this thesis. To be able to test the prototype implementation of the generation component, a component that outputs (specified) MEPs on the basis of information specified in a text file was implemented. In the following, this component is called 'MEP Simulator'. The formulation component was implemented as discussed in Section 4.7.2. In the following evaluations of the generation system, voices from Acapela's[24] Infovox3 synthesis software were used.

Evaluation and Discussion of the Generation Prototype

To show the function of the generation prototype, example inputs that were derived by manually annotating screen-records from experimental data were used. This data was previously collected in the experiments discussed in Chapter 3. According to structure of the VAVETaM system, the MEP Observer component provides such output, as discussed

[24]http://www.acapela-group.com

in Section 4.6.2.[25] The manually annotated MEPs and specifications were input to the prototype of the generation component for testing purposes. The MEP Simulator was used to enable this testing.

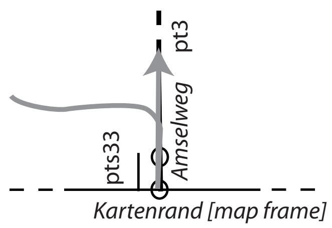


Figure 4.14: Exploration movement a visually impaired participant performed in an experiment.

In order to exemplify the function of the generation prototype, a small part of one of the annotated inputs is detailed in this section.[26] Figure 4.14 visualizes a part of the movement of a visually impaired map user and the corresponding names and identifiers of the objects used for the specification of the MEPs in the VETM knowledge base (the full map is shown in Appendix D). As the figure shows, the map user touches the track pt3, coming from the left. The track is explored for a while with small movements. (This position is remained for a relatively long time, maybe listening to the ongoing utterances.) Then, the map user's exploration movements proceed to the bottom end of the track before following the track upwards. Figure 4.14 shows that the bottom end of the track is conceptualized as distinct track segment, track segment pts33, which is part of the track pt3.

Time in Seconds	**Input to the GVA**
...	...
33.0–54.0	trackMEP(pt3)
54.0–57.0	trackMEP(pt3, [pts33])
57.0–57.8	trackMEP(pt3)
...	...

Table 4.8: Manually categorized MEPs and specifications for the exploration depicted in Figure 4.14.

The annotated MEPs and their specification of this small exploration movement are shown in Table 4.8. The generation component and the formulation and articulation components generate detailed log files. These files indicate which information is selected

[25]As the function of the prototype of the VAVETaM system reported in Section 4.8 shows, an implementation of the MEP Observer component created by Matthias Kerzel outputs specified MEPs as suggested. This shows the technical feasibility of the approach.

[26]Other annotated inputs were also tested; this example is representative of the behavior of the prototype.

from the VETM knowledge base, which preverbal messages are put onto the Agenda, and how utterances are formulated. Based on the log files, I discuss the processes performed by the generation component. The resulting verbal output is shown in Table 4.9.

During the user's long first exploration movement the user focuses on the track pt3. This exploration movement goes on from seconds 33 to 54. It is conceptualized as trackMEP(pt3). The generation component constructs preverbal messages containing all information that is associated with the track pt3 in the VETM knowledge base (the referential net used is shown in Appendix C). The first utterance produced informs the user about the identity of the track. It can be translated to: 'This is Amselweg.' This utterance is of message class 1.1. Then, the user is informed about geometric relations of this track with other tracks (message class 2.3). Information about parallelism with the track pt4 is available in the VETM knowledge base and a corresponding utterance is produced. Subsequently, the user is informed about the extension of the track (message class 2.1); that is, where it ends. Then, information about intersections the track has is uttered (message class 2.2). These are all assisting utterances that are possible given the current exploration situation and the information contained in the VETM knowledge base.

Next, the user moves towards the map frame and the track segment pts33 comes into secondary focus. Thus, a message that informs the user about his or her position on the track segment is formulated and put in front of the (anyways empty) Agenda, resulting in an utterance that can be translated as 'Here, Amselweg is restricted by the map frame'. When the user leaves the track segment pt33, no further assisting utterances are given as the Agenda is able and all information associated with the track pt3 has already been uttered.

Table 4.9: The processes in and the output (German and translated) of the generation (GVA) and formulation components.

...
33.0–54.0 s
MEP Simulator fires trackMEP(pt3)
GVA receives: trackMEP(pt3)
GVA clears Agenda due to MEP change
PVM Construction is able to generate PVMs of class: Identification, Geometric Relation, Extension, Junctions
PVMs Identification, Geometric Relation, Extension, Junctions, are put on the Agenda (0 omitted, no recent articulation according to Verbalization Memory)
Formulation getting Identification PVM for the RefO pt3: the following aspects have been chosen by PVM Construction: name 'Amselweg'
Utterance: "Dies ist der Amselweg." [*"This is Amselweg."*]
Formulation getting Geometric Relation PVM for the RefO pt3: the following aspects have been chosen by the PVM Construction: is_parallel_to with the arguments [pt3, pt4]
...

...
Utterance: "Parallel zu ihm verläuft die Blumenstraße." [*"...which is parallel to Blumenstraße"*]
Formulation getting Extent PVM for the RefO pt3: the following aspects have been chosen by the PVM Construction: predicate has_upper_limit with the arguments [pt3, ptco1]; predicate has_lower_limit with the arguments [pt3, pfr3]
Utterance: "...er mündet nach oben in die Dorfstraße und endet unten am Kartenrand." [*"...it forms a corner with Dorfstraße at the top and at the bottom is restricted by the map frame."*]
Formulation getting Junctions PVM for the RefO pt3: the following aspects have been chosen by the PVM Construction: predicate is_in_track_config with the arguments [pt3, ptco4]
Utterance: "Außerdem hat er eine Kreuzung mit der Hochstraße." [*"Furthermore, the street crosses Hochstraße."*]
54.0–57.0 s
MEP Simulator changes MEP specification to trackMEP(pt3, [pts33])
GVA receives: trackMEP(pt3, [pts33])
GVA detects secondary focus change
PVM Construction is able to generate PVMs of class: Identification
Identification PVM is put at the front of the Agenda (0 omitted, no recent articulation according to Verbalization Memory)
Formulation getting Identification PVM for the RefO pts33
Utterance: "Hier endet der Amselweg am Kartenrand." [*"Here, Amselweg is restricted by the map frame."*]
57.0–57.8 s
MEP Simulator changes MEP specification to trackMEP(pt3)
GVA receives: trackMEP(pt3)
Nothing happens, primary focus not new
...

4.7.6 Prototype Implementation of the Generation Component: Temporal Robustness

As described in the preceding section, given plausible input, the component shows reasonable performance. However, the haptic event detection is a process that has to work in an incremental manner: Haptic exploration events have to be detected while they are still performed and before they are finished. This detection of exploration events should be as timely as possible. Yet, detection should not make errors (i.e., conceptualize exploration situations wrongly) because this would lead to inappropriate utterances. For example, the haptic event detection should not detect a trackMEP when the map user just crosses a track, because this can lead to utterances concerning the track and its surroundings in situations when the track is not in the haptic focus (e.g., 'This is Dorfstraße. ...').

In incremental processing, the time span until a decision about the output is made is crucial for how many wrong decisions (i.e., decisions that need revisions) a detection component makes. Therefore, the longer the period for detection is, the smaller is the amount of wrongly detected exploration situations. In other words, in incremental processing, there is a tradeoff between timeliness and certainty of the correctness of the output (i.e., if it is stable or needs revision at a later stage) (Baumann, Atterer, & Schlangen, 2009). More demands on certainty of the correctness of the output imply a longer time until exploration situations can be detected.

Table 4.10: A part of the original annotation without delay (see Figure 4.15, for the relevant part of the map).

Time in ms	(Specified) MEP
...	...
49090	trackMEP(pt4,[ptco2])
49210	trackMEP(pt4)
49520	trackMEP(pt4, [pts44])
49770	trackMEP(pt4, [ptco5])
49910	trackMEP(pt4)
50050	trackMEP(pt4, [pts46])
...	...

The robustness of the generation component towards different kinds of input was tested. Therefore, input that varied with respect to the timeliness of the detection of MEPs and their specification was fed into the component. The input was fed into the generation prototype with the MEP Simulator. An additional screen-record of one of the experiments was manually annotated (see Chapter 3). In this annotation, each MEP was annotated as soon as possible based on the position of the map user. For example, a specified trackMEP was annotated as soon as a track was touched. As shown in Table 4.10, this process resulted in a large frequency of changes in MEPs and, especially, in their specification. Due to the tradeoff between timeliness and certainty of the correctness in incremental processing, it is likely that an incremental component does not detect MEPs and specifications as soon as they start; that is, an incremental component detects MEPs and specifications after the respective exploration movements have been started. To test for the effects of delay on the output of the prototype of the generation component, delays were systematically introduced in the original annotation.

Four types of delays were introduced: (1) the delay of detecting that an MEP starts, (2) the delay of detecting that an MEP has ended, (3) the delay of detecting secondary foci of a current MEP, and (4) the delay of detecting that secondary foci are obsolete. Multiple delays were introduced because the introduction of only one delay would not have changed the behavior of the system, but rather delayed the start and the end of each MEP and specification equally, resulting in exactly the same behavior as with the original input. Furthermore, detecting that an MEP has started is a much more complicated task

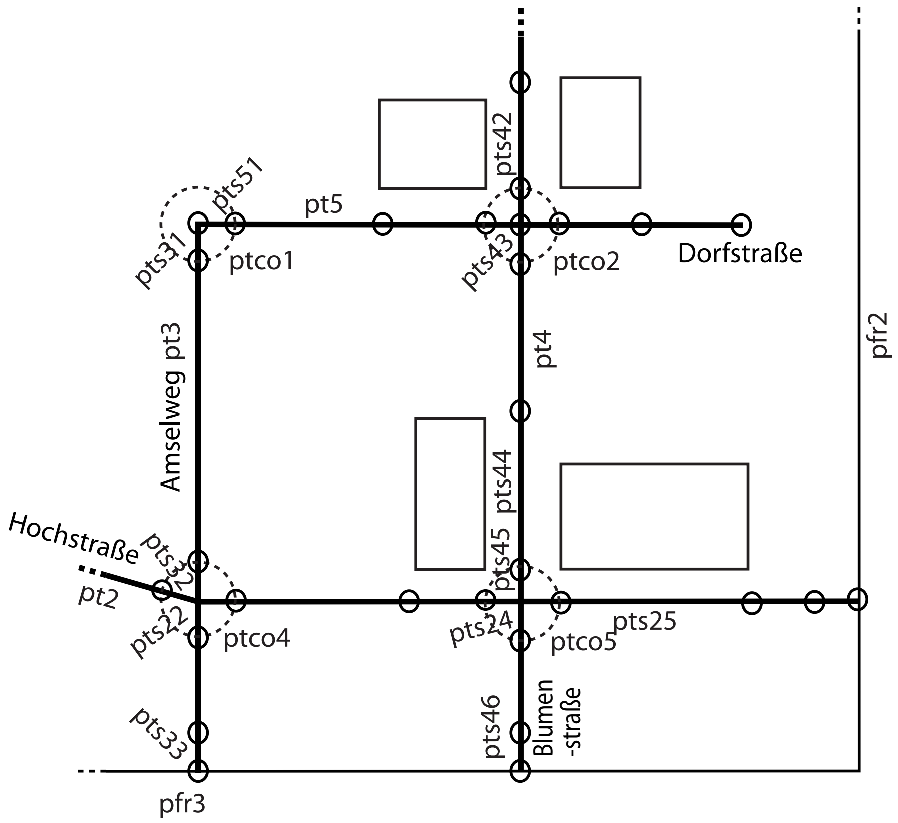


Figure 4.15: The relevant part of the map (the full map is shown in Appendix D).

then detecting that it has ended: The MEP Observer component has to track whether a specific exploration movement (such as following the course of a track or the outline of a landmark) occurred. Rather, to detect that an MEP has stopped, it is sufficient to realize that the interface point is not located on the map entity anymore; that is, in the latter case no analysis of the movements of the interface point has to be made. Basically, the same argument justifies the introduction of separate delays for the detection of secondary foci: It is sufficient to identify that the interface point is located in the part of the map that is associated to the map entity that constitutes the secondary focus or, respectively, that the interface point is not located on it anymore. From the original manually annotated input file, the delayed versions were calculated using custom-built software. Articulation for this evaluation was realized with Acapela Infovox 3 and Chant SpeechKit 6.

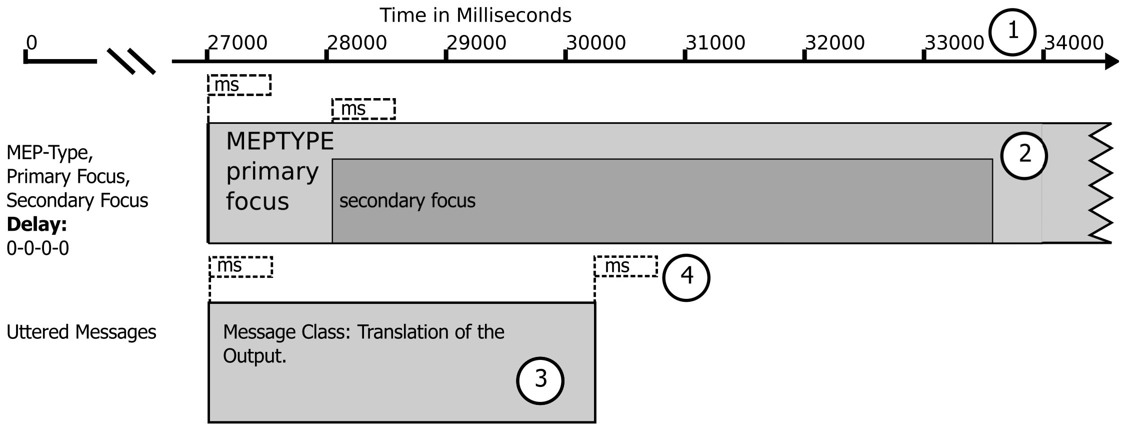


Figure 4.16: Explanation of the visualizations of the input to the generation component and the corresponding output.

The original and delayed inputs were fed into the prototype of the generation component. To show the effect of the delays on the working of the prototype, I visualize the input to the generation component and its output as exemplified in Figure 4.16. At the top of the figure, a time line shows the progression of time (see marker 1). The time from the beginning of the exploration is noted in milliseconds (ms). In the left part of the the figure, under the bold-printed heading 'Delay', are four numbers that indicate whether and how much the orignial input was delayed (in the example figure, the numbers 0-0-0-0 indicate that there was no delay introduced). The numbers represent the delay in milliseconds. The first number indicates the delay introduced until an MEP starts (delay type (1)). The second number indicates the delay until an MEP ends (delay type (2)). The third number indicates the delay of detecting secondary foci (delay type (3)). The fourth number indicates the delay of detecting that secondary foci are obsolete (delay type (4)). Below the time line, grey boxes show the sequence of MEPs and their specification that are the input to the generation component (see marker 2). The zig-zag line to the right of the box indicates that the respective MEP is longer than shown in the figure. The lighter-grey and larger box shows an MEP. The type of the MEP is indicated on the top left of this box, followed by a specification of the primary focus. Within the lighter grey box, smaller and darker-grey boxes indicate that during a certain time period the MEP is specified with secondary foci. The identifier(s) of the respective map entities are noted in the box. Below these two boxes, another grey box indicates the output of the generation component (see marker 3). Within this box, the message class of the output is specified along with a translation of the utterance that is output. The beginnings and the endings of MEPs and utterances that are output are marked with flags with dashed lines (see marker 4). The numbers in this flags are the exact time points in milliseconds.

In Figure 4.17 and Figure 4.18, two small parts of an input file with varying delays and the corresponding output are shown. The visualizations are based on log files the component generates. A visualization of the relevant part of the corresponding map is shown in Figure 4.15. The exploration situation is similar to that discussed in Section 4.14 (yet, the screen record of the movements on which the annotation was based stems from another participant). Figure 4.17 (a) shows the performance of the component from second 26 to second 37 with the original annotation without delay as input. We tested the performance with longer inputs (the full input in all four delays was about 180 seconds), I present only a part of the data to keep the visualization and discussion clear.

When the figure begins, the exploration is conceptualized as trackMEP(pt2). At millisecond 26470, the crossing ptco4 comes in the haptic focus. At millisecond 26734 the utterance of an identification message for the secondary focus is started. The user receives the utterance, which is translated to 'Here is the intersection between Amselweg and Hochstraße'. While this message is uttered, at millisecond 27010, the MEP switches to trackMEP(pt3, [ptco4]). According to the utterance plans, preverbal messages for pt3 that can be constructed from the VETM knowledge base are put on the Agenda after the Agenda is cleared from other information. The identification message for the secondary focus ptco4, which would usually be put in front to the Agenda, is prevented because it was recently sent to the articulation component. At millisecond 27640, the intersection gets out of the haptic focus. Then, starting at millisecond 28050, the map user explores the end of the track, which is represented by the map entity pts33 on the propositional layer of the VETM knowledge base.

As discussed in Section 4.7.4, messages for secondary foci are usually more time critical than messages for primary foci. Therefore, the corresponding identification message is put in front of the Agenda. However, the articulation of the identification message for the intersection is still ongoing and therefore, articulation of the new message cannot be started. At millisecond 29968, the articulation of the message for the intersection ends. We introduced a pause of 600 milliseconds because this results in more natural utterances.[27] At millisecond 30578, the articulation of the next message, which is again an identification message for the secondary focus starts. The articulation starts after

[27]The value of this parameter was adjusted based on informal testing.

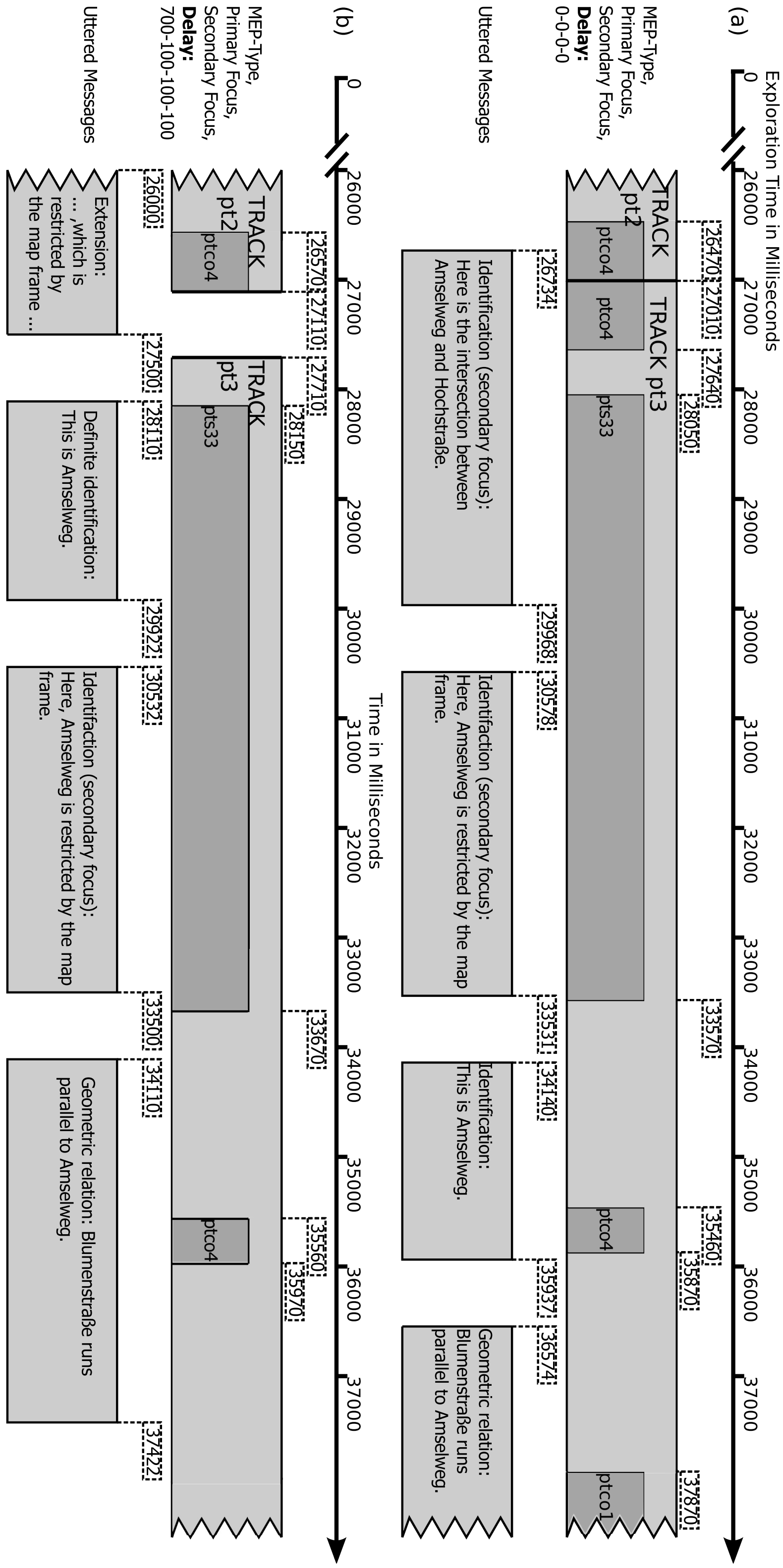


Figure 4.17: Visualization of the performance of the generation component without delay and with delays of 700 ms for the start of MEP detection and 100 ms for the end of MEP detection and for detections of secondary foci.

the pause of 600 milliseconds and additional 10 milliseconds processing time needed for formulation and preparing for articulation. At millisecond 33531, the articulation component finishes the articulation of this message. At millisecond 33570, pts33 gets out of the secondary focus. After the pause of 600 milliseconds and additional processing time of 9 milliseconds, the next message from the Agenda is articulated beginning at millisecond 34140. This message is an identification message for pt3, which informs the user about the name of the track.

Whether this message is needed at this time point is questionable, as the identification message for pts33 presupposes this information. On the other hand, it is possible that the user has passed crossing ptco4 again and, hence, does not know if he or she is still on the track with the name 'Amselweg' or whether the track with that name ends in the crossing. Therefore, it is reasonable to state the utterance to make sure that the user knows that he or she is still on the track called 'Amselweg'. In the case that the MEP Observer component, when fully implemented, reliably detects all secondary foci, reasoning in the Map-Knowledge Reasoner can estimate whether this message should be uttered or not. However, this estimation needs a fully implemented and elaborated Map-Knowledge Reasoner component that works in a multi-modal manner (a part of the VAVETaM system which is presently not existing). Currently, the message is uttered because it is less critical when additional verbal information is given than when important information is left out.

While the identification message for pt3 is uttered, ptco4 comes in the haptic focus again for a very short time period (milliseconds 35460 to 35870). Whether the according utterance is put on the Agenda depends on the setting of the parameters of the memory component (i.e., for how long repetitions of identification messages for secondary foci are prevented).[28] However in the example, during the respective time period, the articulation of another utterance is still ongoing and the secondary focus is obsolete before the articulation is idle. Hence, no message associated to ptco4 is uttered.

As no information is added to the Agenda, the next message that is uttered at millisecond 36574 (37 milliseconds are needed for processing) informs the user about geometric relations that the track has with other track segments, which was put on the Agenda when the trackMEP started, according to the utterance plans.

Figure 4.17 (b) shows the performance of the component when the same input is delayed for 700 milliseconds (delay type (1), delay of start of the MEP), and 100 milliseconds (delay types (2)–(4), delay of end of MEP and recognition of secondary foci start and end). As can be seen at 26000 milliseconss, the utterance of an extension message is ongoing. This differs from Figure 4.17 (a) and is due to the different verbalization histories (stemming from the introduced delay). As Figure 4.17 (b) shows, the trackMEP that is ongoing when the figure starts lasts 100 milliseconds longer according to the delay introduced. From milliseconds 27110 to 27710, no MEP is detected. This gap is caused by the differences in the start delay and the end delay.

[28] Furthermore, as suggested in Lohmann et al. (2011), in this case the utterance could be generally prevented as the map user explores the map very fast. Therefore, it is not very likely that a time critical identification message can be integrated properly by the map user.

At millisecond 27710 the trackMEP specified with pt3 starts. According to the utterance plans, information corresponding to the track pt3 is stored on the Agenda, resulting in the utterance of an identification message for the track at millisecond 28110 (after the 600 milliseconds pause between utterances and 10 milliseconds of processing time). Shortly after the utterance is started, at millisecond 28150, the track segment pts33 comes into the second focus. An identification message for the secondary focus is put in front of the Agenda. The utterance of this message is started at 30532 milliseconds. Later on, at millisecond 34110, a geometric relation message is uttered. This message was put on the Agenda when the trackMEP was detected and, as the MEP is ongoing, is still current and on the Agenda.

Figure 4.18 (a) shows the same input with the delay 1500-500-500-100. As shown in the figure, the messages that are uttered are very similar to that which are uttered in the other two delay conditions. Altogether, utterances are started later in comparison to the other conditions with less delay. This behavior is expected, as utterances can only be started when the corresponding MEP is input to the component, which is delayed in the delay conditions.

Figure 4.18 (b) shows the performance of the component with a delay of 2 seconds for the detection that an MEP starts or ends and 1.8 seconds for the detection that secondary foci are current or obsolete. As can be seen in the figure, the generation, formulation and articulation components still produce output comparable to that produced in the other delay conditions. Note that the formulation of the definite identification message starting at milliseconds 35313 differs from the formulation in the other conditions. This is not due to the translation, but according to the original data (the original German output in this condition is 'Jetzt bist du auf dem Amselweg' and in the other conditions it is 'Das ist der Amselweg'). As described in Section 4.7, the formulation component produces some variation in order to make the utterances more natural.

The performance of the component is not negatively affected by delay introduced in the input, as exemplified by the small example of the performance of the component that was discussed in detail. The components processing results in reasonable output in all delay conditions. Furthermore, even though especially the original input contained a comparable large frequency of changes in MEPs and specification (as exemplified in Table 4.10, as shown, the specification of the MEP changes 6 times in less than a second), the prototype of the component produces reasonable output.

I discussed the performance with a delay for detecting MEPs up to two seconds. We also tried larger delays and the generation component is able to produce plausible output. However, in the situated map-exploration scenario, a delay of 2 seconds is already quite large. Therefore, I did not discuss larger delays here. Especially crucial, however, is the delay of detecting that an MEP has stopped. A delay too large in this case results in the utterance of messages that do not correspond to the map user's exploration situation, which is very confusing and counterproductive. The same holds true for detecting that secondary foci are outdated.

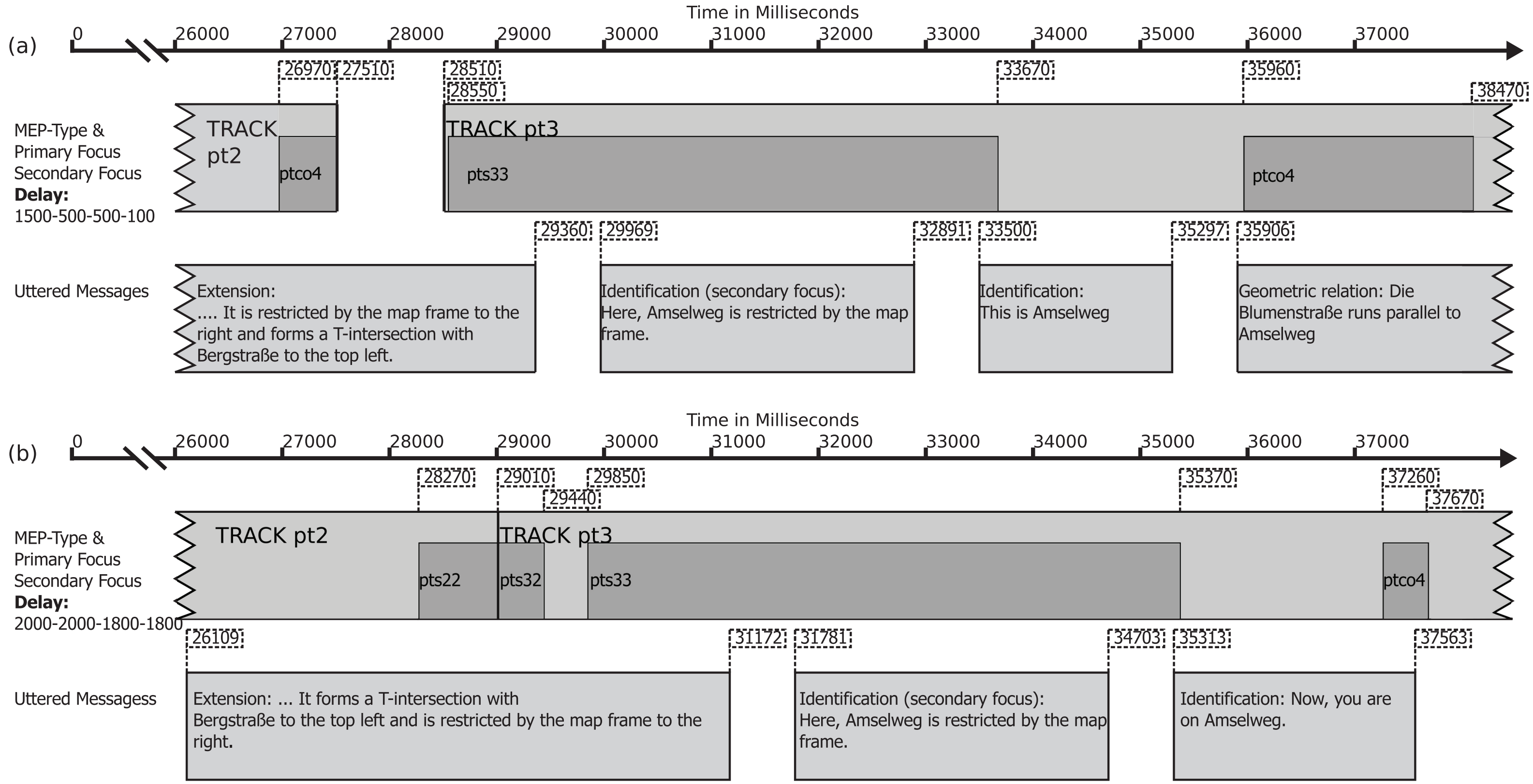


Figure 4.18: Visualization of the performance of the generation component with delays 1500-500-500-100 and with delays 2000-2000-1800-1800.

4.8 The VAVETaM Prototype

As reported in Lohmann, Kerzel, and Habel (2012), we connected the discussed language-generation prototype to a prototype of the MEP Observer component (event detection is discussed in Kerzel & Habel, 2012). Specified MEPs as discussed serve as interface between the event-detection component and the generation component. The system is an early prototype of the VAVETaM system. To test the prototype and to collect impressions for further improvements, we conducted a user study with 13 blindfolded sighted participants (university students, compensated by partial course credit or monetary; mean age: 24.7 years, $SD = 7.2$ years, 9 males).

Participants were trained with the haptic device. Afterwards, they were blindfolded and used the VAVETaM prototype. When they reported that they understood the behavior of the prototype, they were instructed to learn the map such that they would be able to sketch it afterwards. Participants could take as long as they wanted to explore and memorize the map (overall interaction time: $M = 14{:}16\,\text{min}$, $SD = 4{:}46\,\text{min}$; interaction after instruction: $M = 9{:}30\,\text{min}$, $SD = 3{:}42\,\text{min}$). After sketching the map, we assessed agreement to statements on Likert-type scales and interviewed participants on their attitudes towards the VAVETaM prototype and what they considered potential improvements. The discussions were audio-recorded.

See Lohmann, Kerzel, and Habel (2012) for a full discussion of the results. I summarize the results regarding the assisting utterances: The agreement to the statements and the answers in the interview indicate that participants considered it helpful to support virtual tactile maps with natural language. This result is inline with the results discussed in Chapter 3. Furthermore, agreements indicate that the prototype reacts in a comprehensive manner. The timing of assisting utterances is crucial to establish reference. Especially when deictic utterances are generated, the event detection and utterance generation have to be quick enough to allow the map user to establish reference (Lawrence et al., 2009, discuss problems with respect to timing in multi-modal maps). The results indicate that the assisting utterances where started fast enough to establish reference to the respective objects in the virtual tactile map. The actually given assisting utterances were considered helpful. Participants knew what was meant with them. Yet, the synthesis quality of proper nouns was not sufficient and is a candidate for improvement. In the user study, we used Mary TTS for synthesis. In addition, we anticipated that it is problematic that the utterance generation functions sentence-wise—that is, once a sentence is started, it cannot be stopped or changed anymore (in contrast, in the Experiment reported in Chapter 3, the experimenter could stop ongoing utterances by clicking on the start button again). Participants confirmed this anticipation, which indicates that stopping or correcting ongoing utterances is another candidate for improvement. Aside from the potential improvements, both with respect to the assisting utterances and to their timing, participants considered the prototype system usable, comprehensive, and helpful.

4.9 Chapter Summary

In this chapter, I discussed the generation component of the VAVETaM system, which solves the task to generate verbal assisting utterances for virtual tactile maps. The utterances produced are comparable to those tested in the experiments discussed in Chapter 3. I discussed the structure of the generation component of the VAVETaM system. Generally, a component that generates natural language has to decide upon the content to be verbalized and prepare it in a manner appropriate for verbalization. A situated system has to solve this task taking the exploration situation (i.e., the current interaction of the user with the system and the interaction history) into account. It additionally has to solve the task to decide when information should be uttered.

I presented the structure of the VAVETaM system, which extends current approaches towards audio-tactile maps. This system realizes VAVETaMs. It decides upon the content of assisting utterances in a situated manner on the basis of the current exploration situation. In contrast, in current approaches , the output of fixed text is triggered when map objects are touched.

This thesis is mainly concerned with the generation component and its inputs. Therefore, the inner workings of the generation component were discussed in detail. Potential utterances are generated on the basis of utterance plans and stored on the Agenda. The Agenda is an ordered list of preverbal messages, which are representations of the content of utterances. The preverbal messages and the position at which they are stored on the Agenda depends on the exploration situation in which the preverbal message is constructed. An exploration situation is defined by the MEP (i.e., the type of exploratory hand movement performed with the device), the object that is focused upon by this movement, and the verbalization history.

To control assisting utterances, MEPs are detected in the stream of movement data. These MEPs are specified with the objects in the haptic focus when appropriate. It is possible that more than one map entity is in the haptic focus. The conceptualization of the map user's hand movements allows for a representation of such situations: MEPs can be specified with multiple haptic foci. The map entity that is the primary subject of exploration is in the primary focus. Additional map entities are represented as secondary foci.

The generation component needs access to knowledge represented in a manner suitable for verbalization. The VETM knowledge base contains this knowledge. It is based on a hybrid representation with two layers. The spatial-geometric layer stores knowledge in a format suitable to present the map with the haptic device and to conceptualize the map user's hand movements. The propositional layer is central to the function of the generation component: On this layer, knowledge is stored in a manner suitable for assistance generation. The representation of knowledge on the propositional layer is based on referential nets (Habel, 1986). This formalism structures the domain in referential objects. The formalism is especially suitable for assistance generation in the multi-modal scenario, because objects serve as anchors to ground the situated utterances.

To show the technical plausibility of the conception of the generation component, a prototype was implemented and tested. The prototype component was tested against different kinds of inputs and produced reasonable output in all cases.

We were also able to develop a prototype of the VAVETaM system by connecting a prototype of the MEP Observer component (the component segments the stream of data from the device to MEPs) to the prototype of the generation component discussed in this thesis (Lohmann, Kerzel, & Habel, 2012). The system produces utterances for an example map as tested in the empirical evaluation. It solves the tasks to analyze the map user's exploration movements, to select appropriate knowledge for verbalization, to prepare it for verbalization, and to articulate it. We conducted a user study with the VAVETaM prototype. Participants considered the behavior of the prototype comprehensive and the assisting utterances generated helpful.

Overall, it can be concluded that a system that generates assisting utterances as tested in the Wizard-of-Oz-like experiments is technically possible. The proposed approach works well—as shown by the testing of the prototype of the generation component in isolation and in the prototype of the VAVETaM system.

Chapter 5

Summary and Conclusion

In this thesis, I argued that assisting utterances for virtual tactile maps facilitate non-visual access to spatial information about environments. External representations of environments such as maps and online maps are important means to support or enable spatial problem solving. Yet, visual (online) maps are not accessible to visually impaired people. Verbally Assisting Virtual-Environment Tactile Maps (VAVETaMs)—a combination of virtual tactile maps presented with a haptic device and assisting utterances—were proposed as interface to information about spatial environments for visually impaired people. The hypothesis was that VAVETaMs facilitate spatial knowledge acquisition for visually impaired people in comparison to current audio-tactile maps that make less extended use of spatial natural language. Specifically, the following four research questions were addressed:

1. What should be said and when should it be said to support non-visual knowledge acquisition of virtual tactile maps?
2. Which methods should be applied to empirically evaluate the approach?
3. Do assisting utterances facilitate non-visual knowledge acquisition of virtual tactile maps?
4. Can a technical solution be used to generate assisting utterances?

Section 5.1 discusses the use of natural language to support spatial knowledge acquisition and the set of assisting utterances developed, as addressed by research question 1 (the timing of assisting utterances; that is, the "when" part of the research question, is further discussed in Section 5.4). Section 5.2 summarizes the results of the experimental work addressed by research questions 2 and 3. The implications of the results that extend the original purpose to evaluate an interface for visually impaired people are discussed in Section 5.3. Section 5.4 summarizes the results concerning the generation prototype and the VAVETaM prototype, and thus, addresses research question 4. This thesis closes with a discussion of some open issues in Section 5.5.

5.1 Assisting Utterances in VAVETaMs

It has been shown in other research that both natural language in the form of descriptions and (virtual) tactile maps can be used as non-visual access to information about spatial environments (e.g., Taylor & Tversky, 1992a; Espinosa et al., 1998; De Felice et al., 2007). Thus, when these representational means are used uni-modally, both natural language and (virtual) tactile maps allow people to build up a spatial mental model of the environment represented. Yet, knowledge about large and unstructured environments such as a city center cannot be acquired only by descriptions. Exploring tactile maps can be more time-consuming and complicated than reading visual maps (Lawrence, 2011). VAVETaMs combine natural language and virtual tactile maps to facilitate non-visual knowledge acquisition. Among external representations of environments, VAVETaMs are a special type. VAVETaMs combine two independent representational systems, namely natural language and virtual tactile maps. Therefore, VAVETaMs are representationally multi-modal.

In previous approaches towards audio-tactile maps, the use of natural language mostly was restricted to informing the map user about the identity of named objects. While evaluations of such systems show the prospects of multi-modal maps combining a haptic representation and using speech to inform the map user about the names of objects (e.g., Jacobson, 1996; Lahav & Mioduser, 2008b), previous research has not addressed a more extended use of natural language to support knowledge acquisition from (virtual) tactile maps. In general, research on using representationally multi-modal information displays as accessible interfaces to spatial information is sparse. However, it has been shown that descriptions of the local surroundings can support spatial knowledge acquisition when indoor rooms are non-visually explored by direct experience (Giudice, 2004; Giudice et al., 2010). This shows the potential of using natural language to support spatial knowledge acquisition.

In this thesis, assisting utterances for virtual tactile maps were investigated. To answer *research question 1*, a set of potentially helpful assisting utterances was developed. A corpus study was conducted to develop this set on an empirical basis. The set includes utterances that allow for the identification of a map object, such as ‘This is Christuskirche’ or ‘This is the crossing between Hochstraße and Amselweg’. Furthermore, the local surroundings of the map object that is currently explored are described with utterances such as: ‘Above Hochstraße, there are the university and the town hall.’ If appropriate, the configuration of the currently explored map object and other map objects are described: ‘This is Hochstraße. It has two crossings, one with Lärchenweg and one with Bergstraße.’ Which utterances are appropriate depends on the sort of the map object that is currently in the haptic focus of the map user’s exloration movements.

5.2 Evaluation of the Effectiveness of VAVETaMs

To evaluate the appropriateness of the set of utterances, the development of adequate empirical methods was necessary (*research question 2*). A combination of quantitative data collection and qualitative research methods was adopted.

With the empirical methods discussed—in particular with the experimental paradigm developed—it was possible to address *research question 3*. Three experiments were conducted to test the effectiveness of assisting utterances of the set developed. In Experiment 1, blindfolded sighted participants took part. While the experiment showed an increase in the knowledge acquired about landmarks when people receive extended assistance (i.e., assistance corresponding to the set of utterances), there was no increase in the knowledge about tracks. Aside from this unexpected finding, Experiment 1 showed that virtual tactile maps are usable and that verbal assistance can be used in the representationally multi-modal setting to facilitate non-visual knowledge acquisition. Furthermore, Experiment 1 showed that the experimental paradigm developed was applicable and suitable.

Experiment 2 showed that a mixed group of visually impaired and blind participants—the potential user group of VAVETaMs—benefited both in their knowledge about tracks and about potential landmarks. Additionally, interviews and system evaluations showed that participants considered the set of utterances helpful and complete and the virtual tactile maps usable. Experiment 2 shows that verbal assistance as suggested helps visually impaired people to acquire knowledge of virtual tactile maps.

However, not all testing methods show this finding significantly. While the data of the recognition task includes a large unsystematic variance that was explained theoretically (see Section 3.6.3), the non-significant result in the tactile sketch mapping task indicates that the task has less sensitivity to detect differences in spatial knowledge. In an experiment including a reconstruction task with visually impaired people reported by Giudice (2004), a similar constellation of outcomes was found: While the data of visually impaired participants showed a significant effect on a distance-estimation task, it showed no significant effect in a reconstruction task.

The results of Experiment 2 show that reconstruction-based methods for testing visually impaired people's spatial knowledge should be applied carefully—at least without additional training procedures. Possibly, tactile sketch mapping would have been more sensitive with a systematical training of participants. In this thesis, it was shown that the agreement between the configuration questions task and the sketch mapping task is lower when visually impaired people perform the tasks than when sighted people perform the tasks given identical stimuli and comparable tasks (Section 3.6.3). As a guideline, I propose to include verbal tasks that are more familiar to visually impaired people in experimental procedures involving this group of people or to systematically test abilities necessary to perform tasks such as tactile sketch mapping.

An additional experiment—Experiment 3—was conducted to investigate the difference between the two groups of participants. Sighted participants explored maps with only tracks and without potential landmarks. To study whether verbal assistance helps at all, participants learned virtual tactile maps without verbal information in addition to the conditions with verbal assistance. All results show that people acquire more knowledge when they receive assistance than when they do not. With respect to the comparison of conditions with verbal assistance, the different testing methods do not all support the same finding. The configuration questions show an increase between the simple-assistance condition and the extended-assistance condition. In contrast, the sketch maps that the participants produced do not indicate improved knowledge acquisition.

The reasons for the different results in the experiments with visually impaired and sighted participants could not be fully explained. In the interviews, visually impaired participants reported that information about tracks is important because it is necessary in navigation. Possibly, visually impaired people are also less interested in information about potential landmarks of the type given in the experimental maps, because they use other landmarks in navigation. Therefore, it is possible that they focus more on tracks than sighted people who can get more information (e.g., about the type of an intersection) during navigation. However, this result does not explain why sketch mapping does not indicate an improvement of sighted people in their knowledge about tracks when they explore track-only maps.

A potential explanation for the results is that spatial knowledge can be represented in different forms and integrated to different degrees (as suggested by earlier evidence discussed in Section 2.5). Interpreting the results of Experiment 3, differences between visually impaired and sighted people with respect to the integration process are a likely candidate for explaining the differences in the results of Experiment 1 and 2, potentially in addition to or governed by a different focus of attention (see Sections 3.7.1 and 3.7.5). It is possible that visually impaired people integrate the representationally multi-modal information to a coherent spatial mental model that preserves spatial (metric) information in an analog fashion to a larger degree than sighted people do.

Altogether, the empirical evidence shows that the set of assisting utterances developed facilitates non-visual knowledge acquisition of virtual tactile maps, but differently for different user groups. The usability of VAVETaMs is further supported by the fact that all 50 participants in the experiments except one were able to pass the test that concluded training. Furthermore, in the user study with the VAVETaM prototype, additional 13 participants were able to interact with VAVETaMs without reporting problems. This shows that the non-visual interaction with VAVETaMs is easy to learn for both sighted and visually impaired people. Furthermore, the results of Experiment 2 show that the knowledge acquisition of the potential user group is facilitated in general by assisting utterances.

The results of the three experiments highlight the need to test interfaces for a special user group with that particular user group. Furthermore, more research is needed to understand how humans process representationally multi-modal information.

5.3 Implications of the Results for Non-Visual Information Processing and Presentation

The results of the empirical evaluation of VAVETaMs are interesting in a broader scope than with respect to human computer interaction and the development of accessible interfaces. Although limited in the generalizability due to the small sample size, the results do not indicate a difference between congenitally blind people and people with a later onset of visual impairment. They clearly show that the mixed group of participants was able to build up internal survey-type representations of space. This finding is in line with other empirical evidence (e.g., Giudice, 2004; Lawrence, 2011). Claims in the literature that blind people are limited to underdeveloped route representations are not supported by the data (see Section 2.4.4 for a discussion of spatial abilities of visually impaired people).

In the experimental conditions with extended assistance, visually impaired participants gained more knowledge about tracks than sighted participants: The data show a significant knowledge increase in comparison to simple assistance for visually impaired people and in contrast, statistically significant equality between conditions for sighted people.[1] With extended assistance, visually impaired participants acquired more knowledge about tracks than sighted people.

The results point towards the possibility that congenitally blind and visually impaired people process representationally multi-modal information about space differently than sighted people. The validity of this finding for congenitally blind people has to be tested for significance with a larger number of participants. In Experiment 2, only three congenitally blind people participated. If this result is valid, it is evidence for the difference theory and speaks against the inefficiency theory (for a discussion of these theories see Section 2.4.4). The inefficiency theory claims that blind people make use of the same representations like sighted people but less effectively and, thus, can only reach spatial performance of sighted people and not outperform them. The results indicate that (congenitally) blind people are able to integrate representationally multi-modal information about tracks better than sighted people. This means that blind people would outperform sighted people at least in some aspects of spatial cognition. Such a result would not be in line with the inefficiency theory.

Furthermore, the results of the experiments show the potential of representationally multi-modal information displays in general and in particular for non-visual communication of information about spatial environments. Humans frequently acquire knowledge on the basis of external representations. These representations need to be perceived and understood. Usually, information about spatial environments is perceived visually. Yet, the results of the experiments reported show that the non-visual acquisition of knowledge is effective, as well.

In particular in external representations that aim to substitute vision as sensory channel and communicate spatial information, representationally multi-modal information displays should be considered. While past research on the processing of multi-modal constellations by visually impaired people has mostly concerned sensory multi-modality (see Pasqualotto & Proulx, 2012, for a review), research on non-visual processing of representationally multi-modal information and the use of representationally multi-modal information displays is sparse (but cf. Acartürk, 2009, for an investigation how visual representationally multi-modal graphs are processed by humans). Especially with respect to non-visual communication of information, representational multi-modality has prospects not only with respect to the domain of urban-area maps.

Aside from maps, graphs are important external representations of information frequently used by humans. Graphs can also be augmented with verbally assisting descriptions and, thus, made accessible to visually impaired people (as proposed by Acartürk,

[1] Note that visually impaired participants did not outperform sighted participants in general but just with respect to the knowledge acquired about tracks in the extended-assistance condition.

2009). Graphs are spatial external representations in which the spatial aspect of the representation (i.e., the analog aspect of the representation) is important to understand them. Graphs share this property with maps. Like maps, graphs are important tools for information communication and problem solving, which are complicated to perceive non-visually.

5.4 Evaluation of the Technical Possibility of VAVETaMs

Additionally to the empirical evaluation of the effectiveness of VAVETaMs, a technical solution was developed and evaluated. This part of the work addresses the technical possibility of automatic assistance generation (*research question 4*). The VAVETaM system extends previous approaches towards audio-tactile maps with respect to how the output is controlled and how output is generated. In previous approaches, the control of verbal output was trigger-based; that is, output is started when a map object is touched. In contrast, the VAVETaM system works on the basis of incrementally detected events in the map user's explorative hand movements. This allows for a more versatile control of verbal assistance. Additionally, output is generated from a semantic knowledge base, which extends earlier approaches towards audio-tactile maps in which output was mostly generated on the basis of fixed, canned texts. To base the generation on a semantic knowledge base allows for the generation of flexible output, which is necessary in the situated scenario.

A structure for the overall VAVETaM system was discussed in Section 4.4. This thesis focussed on the generation component of the VAVETaM system. Furthermore, the inputs to the generation component were discussed. The generation of assistance while the user is exploring virtual tactile maps is a case of situated language generation, which results in specific demands such as timely interaction of haptic exploration and verbal output. A major demand is that the component has to decide when content has to be uttered. In the generation component presented, the control of the output is based on a representation of exploration events that the map user performs. Event-based control of the output has also been used in earlier systems that have the goal to describe dynamic scenes (André et al., 1988; Guhe et al., 2004). The generation is based on knowledge of which information is potentially helpful to verbalize in an exploration situation. Semantic representations of utterances—preverbal messages (Levelt, 1989)—are generated on the basis of this knowledge and the semantic knowledge base. The preverbal messages are stored on the Agenda, which enables an incremental planning of the next utterances.

A prototype of the generation component was successfully tested for plausible output and for its robustness against different types of input. As reported in Lohmann, Kerzel, and Habel (2012), we were also able to connect the generation prototype to a prototype of the component detecting exploration events—resulting in a prototype of the complete VAVETaM system. The interface between the two components are representations of current exploration events, as discussed in this thesis. The functioning of the prototype was confirmed in a user study (Lohmann, Kerzel, & Habel, 2012) of which the relevant parts were presented in this thesis.

Altogether, the functioning of the prototype of the generation component shows that situated assistance generation for VAVETaMs is technically possible and that the approach developed is appropriate to solve the task.

5.5 Open Issues

Some open issues were identified in the empirical work on VAVETaMs and the VAVETaM prototype. These issues are related to the development of the generation component and to human processing of representationally multi-modal information.

In the direction of the development of the generation component, the user study with the prototype of the VAVETaM system indicated some improvements regarding language generation. Most importantly, the possibility to change ongoing utterances when the exploration situation changes was identified as a potential improvement (see Section 4.8). Currently, the generation works sentence-wise—that is, once articulation of a sentence is started, it cannot be stopped or changed anymore. For instance, when map users move to other map entities during their exploration movements shortly after articulation of a long sentence was started, it is not possible to stop or change this sentence. Recently, incremental synthesis has been developed, which allows to change ongoing utterances (Baumann & Schlangen, 2012). Incremental synthesis enables to extend the generation part (including formulation and articulation of utterances) of the VAVETaM system such that it is possible to change or stop ongoing sentences. Such a change requires a fully incremental architecture and feedback from the synthesis to the formulation and the generation component, because information on how far articulation of an utterance has proceeded is needed to enable re-planning in the generation and formulation components. Furthermore, extensive knowledge about when to stop or change utterances is needed. Under which circumstances (e.g., how much of the utterance has already been articulated) utterances should be stopped or changed and whether linguistic markers such as introducing pauses, interjections, or hesitation sounds (such as 'ähms') should be used to indicate that an utterance has been changed are open issues for further research.

With respect to human-processing of representationally multi-modal information, differences in the processing of such information were found between sighted and visually impaired people, as discussed earlier in this chapter. However, only initial steps towards an explanation for the difference could be made. While the results show problems with using classical behavioral testing methods such as tactile sketch mapping with blind and visually impaired people, at least with respect to their sensitivity, other methods can be used to investigate spatial mental models of visually impaired people. Mental scanning paradigms can be used to investigate whether spatial mental models have analog properties. Such paradigms have also successfully been used to test whether spatial knowledge of blind people has analog properties (e.g., Röder & Rösler, 1998; for a comprehensive review of applications, see Cattaneo & Vecchi, 2011). If knowledge about environments is analog, it preserves metric spatial properties of the environment. Therefore, it also needs to be integrated to a large degree. Scanning paradigms are candidates to test whether the spatial knowledge of visually impaired and sighted people encodes space in an analog way. Mental scanning paradigms can be used to test the knowledge acquired non-visually from representationally multi-modal sources of information about environments.

As discussed, Experiment 3 showed that sighted people significantly increased in some tests of spatial knowledge when they learned track-only VAVETaMs, but not in others (Section 3.7.3). A possible explanation for the pattern in the outcomes is based on the following observation: Participants improved in all tests in which no strong integration of spatial knowledge to a coherent mental model is necessary. If it can be proven that blind people integrate the knowledge perceived to a coherent spatial mental model, but sighted people do not, this would provide strong support to the hypothesis of processing differences. Because such a result would show that blind people do not rely on the same internal representations and processes as sighted people just used less effective, this would strongly argue against the ineffectiveness theory. If spatial knowledge is represented in an analog manner, it also needs to be coherent and integrated to a high degree: An analog representation preserves relevant constraints of the represented world and, consequently, does not allow for inconsistent knowledge (see Section 2.2.1). The empirical results reported indicate differences in the processing of representationally multi-modal spatial information between visually impaired and sighted people. The successful applications reported in the literature indicate that mental scanning paradigms can be used to further investigate these potential processing differences.

Appendix A

Excerpt of the Corpus of Human Assistances

In the following, an excerpt of the corpus collected is presented. The table shows data from five different assistants. In the column labeled 'message class', the classification of the utterance according to the message classes discussed in Section 3.2 is shown. A questionnaire indicates that a clear classification was not possible. If an utterance was a comment without the purpose to help the map explorer, no message class is assigned.

No.	Utterance	Split Utterance Analyzed	Message Class
1	Hmm, geht das schnell		
2	Ähm, so du befindest dich jetzt am		
3	Ja du, du bist zu schnell, da kommt man kaum mit		
4	So du befindest dich nun am östlichen, westlichen Kartenrand, jetzt am südlichen Kartenrand	Das ist der östliche Kartenrand	1.1
5	"	Das ist der westliche Kartenrand	1.1
6	"	Das ist der südliche Kartenrand	1.1
7	Bist du am westlichen Kartenrand ... und ...	Das ist der westliche Kartenrand	1.1
8	Jetzt hast, bist du auf den Kirchweg gekommen	Das ist der Kirchweg	1.1
9	Der Kirchweg verläuft diagonal durch die Karte	Der Kirchweg verläuft diagonal durch die Karte	2.3
10	Bis, bis er auf die Dorfstraße trifft	Der Kirchweg trifft auf die Dorfstraße	2.2
11	Du bist jetzt auf der Dorfstraße	Das ist die Dorfstraße	1.1

No.	Utterance	Split Utterance Analyzed	Message Class
12	Die Dorfstraße läuft ähh auch diagonal und trifft mit einem 90-Grad-Winkel auf den Kirchweg - ungefähr	Die Dorfstraße verläuft diagonal durch die Karte	2.3
13	”	Die Dorfstraße trifft ungefähr mit einem 90-Grad-Winkel auf den Kirchweg	2.2/2.3 (?)
14	So genau		
15	Bist' auf bei der T-Kreuzung von der Dorfstraße	Du bist auf der T-Kreuzung von Dorfstraße und Marktstraße	1.1
16	Jetzt bist du auf die Marktstraße abgebogen	Du bist auf die Marktstraße abgebogen	1.1
17	Du bist wieder auf der Dorfstraße	Das ist (wieder) die Dorfstraße	1.1
18	Die Dorfstraße bildet den südlichen Rand vom Dorfplatz	Die Dorfstraße bildet den südlichen Rand vom Dorfplatz	4.1
19	Der Kirchweg und Dorfstraße schließen den Dorfplatz ein, auf dem eine Kirche steht	Der Kirchweg und die Dorfstraße schließen den Dorfplatz ein	4.1
20	”	Auf dem Dorfplatz steht eine Kirche	3.2
21	Du explorierst nun gerade die Kirche, die auf dem Dorfplatz steht	Das ist die Kirche	1.1
22	”	Die Kirche steht auf dem Dorfplatz	3.2
23	Du bist wieder auf dem Kirchweg	Du bist wieder auf dem Kirchweg	1.1
24	Und der Kirche, die auf dem Dorfplatz steht	Das ist wieder die Kirche	1.1
25	”	Die Kirche steht auf dem Dorfplatz	3.2
26	Der Dorfplatz ist auf der westlichen Seite der Karte und bildet ein Dreieck, eingeschlossen von Kirchweg und Dorfstraße	Der Dorfplatz ist im westlichen Bereich der Karte	3.1
27	”	Der Dorfplatz bildet ein Dreieck	4.2

No.	Utterance	Split Utterance Analyzed	Message Class
28	”	Der Dorfplatz ist eingeschlossen von Kirchweg und Dorfstraße	4.1
29	Jetzt befindest du dich gerade in der Hochstraße und der neuen Gasse und der alten Gasse, die von der Marktstraße aus . . .	Das ist die Hochstraße	1.1
30	”	Das ist die neue Gasse	1.1
31	”	Das ist die alte Gasse	1.1
32	Die Marktstraße führt auf den Dorfplatz	Die Marktstraße führt auf die Dorfstraße	2.2
33	Von der Marktstraße abgehen tut die alte Gasse auf der du gerade warst	Du warst gerade auf der alten Gasse	1.1
34	”	Die alte Gasse geht von der Marktstraße ab	2.2
35	Das ist die Marktstraße, genau	Das ist die Marktstraße	1.1
36	Du bist wieder auf der Dorf...	Das ist (wieder) die Dorfstraße	1.1
37	Das ist, genau Dorfstraße Zu deiner link-li-links von der Dorfstraße ist ja die Marktstraße, genau das ist sie	Links von der Dorfstraße ist die Marktstraße	1.2
38	”	Das ist die Marktstraße	1.1
39	Darunter ist die alte Gasse, die über die neue Gasse in die Hochstraße führt, die in einer Sackgasse am südlichen Kartenrand endet	Unter der Marktstraße ist die alte Gasse	1.3
40	”	Die alte Gasse führt über die neue Gasse in die Hochstraße	2.2
41	”	Die Hochstraße endet in einer Sackgasse	2.1
42	”	Die Sackgasse, in der die Hochstraße endet, liegt am südlichen Kartenrand	3.1

No.	Utterance	Split Utterance Analyzed	Message Class
43	Ähm, du bist jetzt am südlichen Kartenrand und äh und der neuen Gasse	Das ist der südliche Kartenrand	1.1
44	"	Das ist die neue Gasse	1.1
45	Das ist die Schule	Das ist die Schule	1.1
46	Die Schule ist -äh- liegt zwisch.. im äh Dreieck von der Hochstraße und der neuen Gasse, sagen wir mal	Hochstraße, Kartenrand und neue Gasse bilden ein Dreieck	4.2
47	"	Die Schule liegt im Dreieck, das Hochstraße und neue Gasse bilden	3.2
48	Jetzt bist - äh - jetzt umrundest du wieder den Dorfplatz		
49	Auf der Dorfstraße	Das ist (wieder) die Dorfstraße	1.1
50	Und nun bist du wieder, gehst du wieder auf der Marktstraße, die von der Dorfstraße aus abgeht und in einer Sackgasse links endet und eine T-Kreuzung zur alten Gasse besitzt	Das ist (wieder) die Marktstraße	1.1
51	"	Die Marktstraße geht von der Dorfstraße ab	2.2
52	"	Die Marktstraße endet in einer Sackgasse links	2.1
53	"	Die Marktstraße und die alte Gasse treffen sich in einer T-Kreuzung	2.2
54	Das ist der west- was war das - der westliche Kartenrand und die Schule, die am südlichen Kartenrand liegt	Das ist der westliche Kartenrand	1.1
55	"	Das ist die Schule	1.1
56	"	Die Schule liegt am südlichen Kartenrand	1.3
57	Die flankiert ist von der neuen Gasse und der Hochstraße	Die Schule liegt zwischen der neuen Gasse und der Hochstraße	1.2

No.	Utterance	Split Utterance Analyzed	Message Class
58	Gerade bist du auf der Dorfs-traße	Das ist die Dorfstraße	1.1
59	Jetzt gehst du über den Dorf-platz	Das ist der Dorfplatz	1.1
60	Und bist in der Kirche	Das ist die Kirche	1.1
61	Jetzt bist du wieder weg-gerutscht		
62	Du gehst gerade den Kirchweg längs	Du folgst dem Kirchweg	1.1
63	Du bist am Rand der Karte	Das ist der Kartenrand	1.1
64	Jetzt explorierst du den Rand der Karte	Das ist der Kartenrand	1.1
65	Das ist kein Weg		
66	Da ist die Karte zu Ende	Das ist das Ende der Karte	1.1
67	Obere linke Ecke	Das ist die Ecke oben links	1.1
68	Jetzt gehst du gerade von links oben nach rechts oben	Jetzt gehst du gerade von links oben nach rechts oben	
69	Jetzt bist du wieder auf der Dorfstraße	Das ist die Dorfstraße	1.1
70	Da fängt die Straße an	Da fängt die Dorfstraße an	2.1
71	Jetzt gehst du darüber	Jetzt gehst du über die Dorfs-traße drüber	1.1
72	Jetzt bist du bei der rechten oberen Ecke	Das ist die rechte obere Ecke	1.1
73	Das ist der Kirchweg	Das ist der Kirchweg	1.1
74	Das ist die Dorfstraße	Das ist die Dorfstraße	1.1
75	Jetzt bist du abgebogen in die Marktstraße	Du bist jetzt in die Markt-straße abgebogen	1.1
76	Da bist du oben rechts		
77	Der Kirchweg	Das ist der Kirchweg	1.1
78	Da ist eine Kreuzung zur Dorfstraße	Der Kirchweg hat eine Kreuzung mit der Dorfstraße	2.2
79	Das ist der Kirchweg	Das ist der Kirchweg	1.1
80	Kreuzung Dorfstraße Kirch-weg, hattest du gerade	Das ist die Kreuzung von Dorfstraße und Kirchweg	1.1
81	Das ist wieder der Kirchweg	Das ist der Kirchweg	1.1
82	Dorfstraße	Das ist die Dorfstraße	1.1
83	Doch das ist ein Platz, das ist der Dorfplatz	Das ist der Dorfplatz	1.1

No.	Utterance	Split Utterance Analyzed	Message Class
84	Auf dem Dorfplatz steht eine Kirche	Auf dem Dorfplatz steht eine Kirche	3.2
85	Das ist die Kirche	Das ist die Kirche	1.1
86	Das ist die Landmarke		
87	Da bist du wieder auf der Dorfstraße zurück	Das ist die Dorfstraße	1.1
88	Du hast den Dorfplatz überquert	Du hast den Dorfplatz überquert	
89	Das ist jetzt die Marktstraße	Das ist die Marktstraße	1.1
90	Das ist die Kreuzung Dorfstraße Marktstraße	Das ist die Kreuzung von Dorfstraße und Marktstraße	1.1
91	Dorfstraße, Marktstraße	Das ist die Dorfstraße	1.1
92	”	Das ist die Marktstraße	1.1
93	Jetzt bist du wieder oben an der Karte und dort ist die Dorfstraße	Du bist im oberen Bereich der Karte	3.1
94	”	Die Dorfstraße befindet sich im oberen Bereich der Karte	3.1
95	Ecke Neue Gasse, Hofstraße Alte Gasse	Das ist die Ecke Neue Gasse Hochstraße	1.1
96	”	Das ist die Alte Gasse	1.1
97	Alte Gasse	Das ist die Alte Gasse	1.1
98	Marktstraße	Das ist die Marktstraße	1.1
99	Dorfstraße	Das ist die Dorfstraße	1.1
100	Das ist die Dorfstraße	Das ist die Dorfstraße	1.1
101	Dorfstraße	Das ist die Dorfstraße	1.1
102	Ecke Marktstraße, Dorfstraße	Das ist die Ecke Marktstraße Dorfstraße	1.1
103	Ecke alte Gasse, Marktstraße	Das ist die Ecke Alte Gasse Marktstraße	1.1
104	Alte Gasse	Das ist die Alte Gasse	1.1
105	Neue Gasse	Das ist die Neue Gasse	1.1
106	Ecke Hochstraße, neue Gasse	Das ist die Ecke Hochstraße Neue Gasse	1.1
107	Hochstraße	Das ist die Ecke Hochstraße	1.1
108	Du bist jetzt in der Schule	Das ist die Schule	1.1
109	Da bist du wieder ganz oben in der Dorfstraße	Du bist ganz oben in der Dorfstraße	?

No.	Utterance	Split Utterance Analyzed	Message Class
110	Wenn du bei der Dorfstraße abgehst, ist ein System aus orthogonalen Straßen	Von der Dorfstraße geht ein System aus orthogonalen Straßen ab	?
111	Hier horizontal die Marktstraße	Hier verläuft die Marktstraße horizontal	2.3
112	Das ist eine Sackgasse	Das ist eine Sackgasse	1.1
113	Die Marktstraße ist eine Sackgasse	Die Marktstraße endet in einer Sackgasse	2.1
114	Jetzt bist du wieder am Ende der Karte	Das ist das Ende der Karte	1.1
115	Sackgasse Marktstraße	Hier endet die Marktstraße in einer Sackgasse	2.1
116	Links abbiegend von der Marktstraße ist die alte Gasse	Links biegt die Marktstraße in die alte Gasse ein	2.2
117	Parallel zur Marktstraße ist die Neue Gasse	Marktstraße und Neue Gasse sind parallel	2.3
118	Da ist wieder die Hochstraße	Das ist die Hochstraße	1.1
119	Wo du gerade warst (ja genau da) das ist eine kleine Sackgasse	Das ist eine Sackgasse	1.1
120	Neue Gasse	Das ist die Neue Gasse	1.1
121	Hier geht es ab in die Buchstraße	Hier geht die Hochstraße ab	1.1
122	Und dann rechts davon ist die Schule	Rechts von der Hochstraße ist die Schule	1.2
123	Nana, also dafür musst du über so einen Platz gehen	Um die Schule zu erreichen musst du über einen Platz gehen	5
124	Marktstraße , alte Gasse	Das ist die Marktstraße	1.1
125	”	Das ist die Alte Gasse	1.1
126	Kirche	Das ist die Kirche	1.1
127	Genau, in der Kirche	Das ist die Kirche	1.1
128	Das ist der Dorfplatz	Das ist der Dorfplatz	1.1
129	Hier ist die Dorfstraße	Das ist die Dorfstraße	1.1
130	Das ist der Kartenrand	Das ist der Kartenrand	1.1
131	Das ist der Kirchweg	Das ist der Kirchweg	1.1
132	Die Dorfstraße	Das ist die Dorfstraße	1.1
133	Das ist der Kartenrand der rechte	Das ist der rechte Kartenrand	1.1
134	Das ist der obere Kartenrand	Das ist der obere Kartenrand	1.1

No.	Utterance	Split Utterance Analyzed	Message Class
135	Das ist die Dorfstraße und die Marktstraße	Das ist die Dorfstraße	1.1
136	”	Das ist die Marktstraße	1.1
137	Das ist die Dorfstraße	Das ist die Dorfstraße	1.1
138	Dorfstraße	Das ist die Dorfstraße	1.1
139	Das war die Ecke Dorfstraße Marktstraße	Das war die Ecke Dorfstraße Marktstraße	1.1
140	Die Marktstraße geht in einer T-Kreuzung in die Dorfstraße über	Die Marktstraße geht in einer T-Kreuzung in die Dorfstraße über	2.2
141	Und das ist die Ecke Dorfstraße Kirchweg	Das ist die Ecke Dorfstraße Kirchweg	1.1
142	Die in einer T-Kreuzung nach rechts abgeht, die Marktstraße geht nach links ab von der Dorfstraße	Die Marktstraße geht nach links ab von der Dorfstraße	2.2
143	Das ist der obere Kartenrand	Das ist der obere Kartenrand	1.1
144	Dorfstraße	Das ist die Dorfstraße	1.1
145	Das obere Ende der Dorfstraße ist begrenzt durch den Kartenrand	Das obere Ende der Dorfstraße ist begrenzt durch den Kartenrand	2.1
146	Das ist die Ecke Dorfstraße, Kirchweg	Das ist die Ecke Dorfstraße Kirchweg	1.1
147	Der Kirchweg geht nach rechts ab, jetzt	Der Kirchweg geht jetzt nach rechts ab	1.1
148	Dorfstraße	Das ist die Dorfstraße	1.1
149	und begrenzt durch den oberen Kartenrand	Die Dorfstraße wird begrenzt durch den oberen Kartenrand	2.1
150	Das ist die Dorfstraße	Das ist die Dorfstraße	1.1
151	Jetzt bist du in der Marktstraße	Das ist die Marktstraße	1.1
152	Die Marktstraße endet links in einer Sackgasse	Die Marktstraße endet links in einer Sackgasse	2.1
153	Du bist über die alte Gasse in die neue Gasse gekommen	Du bist entlang der Alten Gasse in die Neue Gasse gekommen	1.1
154	Das ist die neue Gasse die rechts in einer Sackgasse endet und links ebenfalls in einer Sackgasse endet	Das ist die Neue Gasse	1.1

No.	Utterance	Split Utterance Analyzed	Message Class
155	”	Die neue Gasse endet rechts in einer Sackgasse	2.1
156	”	Die neue Gasse endet links in einer Sackgasse	2.1
157	Das ist Hochstraße die durch den unteren Kartenrand begrenzt wird	Das ist die Hochstraße	1.1
158	”	Die Hochstraße wird durch den unteren Kartenrand begrenzt	2.1
159	Die neue Gasse	Das ist die Neue Gasse	1.1
160	Und die Hochstraße	Das ist die Hochstraße	1.1
161	Jetzt bist du in der Schule	Das ist die Schule	1.1
162	Jetzt bist du in der Hochstraße	Das ist die Hochstraße	1.1
163	Du hast jetzt alles auf der Karte bis auf die Kirche exploriert	Du hast jetzt alles in der Karte bis auf die Kirche exploriert	5
164	Die Kirche ist auf den Dorfplatz	Die Kirche ist auf dem Dorfplatz	3.2
165	Jetzt bist du wieder auf der Karte	Jetzt bist Du wieder auf der Karte	5
166	Der im rechten Bereich der Karte ist	Der Dorfplatz ist im rechten Bereich der Karte	3.1
167	Und von Kirchweg und Dorfstraße begrenzt wird	Der Dorfplatz wird von Kirchweg und Dorfstraße begrenzt	4.2
168	Das ist die Dorfstraße	Das ist die Dorfstraße	1.1
169	Und das ist der Teil der Dorfstraße, der den Dorfplatz begrenzt	Das ist der Teil der Dorfstraße, der den Dorfplatz begrenzt	1.1
170	Jetzt bist du auf dem Platz über dem Kirchweg	Das ist der Platz über dem Kirchweg	1.1
171	Jetzt bist du ganz oben in der Karte	Jetzt bist du ganz oben in der Karte	3.1
172	Und Dorfstraße	Das ist die Dorfstraße	1.1
173	Die Marktstraße	Das ist die Marktstraße	1.1
174	Du bist jetzt an der nördlichen rechten Ecke der Karte	Das ist die Ecke im Norden rechts	1.1

No.	Utterance	Split Utterance Analyzed	Message Class
175	Du bist jetzt auf der Tannenstraße. Die Tannenstraße verläuft von links nach rechts und ist die nördlichste der Straßen	Das ist die Tannenstraße	1.1
176	”	Die Tannenstraße verläuft von links nach rechts	2.3
177	”	Die Tannenstraße ist die nördlichste der Straßen	3.3
178	Die Tannenstraße endet rechts in einer Sackgasse, du bist gerade an dieser Sackgasse angelangt	Die Tannenstraße endet rechts in einer Sackgasse	2.1
179	”	Das ist die Sackgasse, in der die Tannenstraße rechts endet	1.1
180	Das ist eine T-Kreuzung die die Tannenstraße mit der Hansastraße verbindet.	Das ist die T-Kreuzung von Tannenstraße und Hansastraße	1.1
181	Die Hansastraße verläuft von Norden nach Süden und ist die am weitesten rechts gelegene Straße die von Norden nach Süden verläuft	Die Hansastraße verläuft von Norden nach Süden	2.3
182	”	Die Hansastraße ist die am weitesten rechts gelegene Straße, die von Norden nach Süden verläuft	3.3
183	Das ist das untere Ende der Hansastraße	Das ist das untere Ende der Hansastraße	1.1
184	Ecke zur Bahnhofstraße	Das ist die Ecke Hansastraße Bahnhofsstraße	1.1
185	Die Bahnhofstraße verläuft von links nach rechts und ist die am weitesten südlich gelegene Straße	Die Bahnhofsstraße verläuft von links nach rechts	2.3
186	”	Die Bahnhofsstraße ist die am weitesten südlich gelegene Straße	3.3
187	Das ist die Schulstraße	Das ist die Schulstraße	1.1

No.	Utterance	Split Utterance Analyzed	Message Class
188	Die Schulstraße ist mit einer T-Kreuzung mit der Bahnhofstraße verbunden	Die Schulstraße ist in einer T-Kreuzung mit der Bahnhofstraße verbunden	2.2
189	Und führt von Norden nach Süden	Die Schulstraße verläuft von Norden nach Süden	2.3
190	Das ist die T-Kreuzung zwischen der Schulstraße und der Bahnhofstraße	Das ist die T-Kreuzung von Schulstraße und Bahnhofstraße	1.1
191	Jetzt bist du wieder auf der Bahnhofstraße	Das ist die Bahnhofstraße	1.1
192	Ecke Bahnhofstraße, Hansastraße	Das ist die Ecke Hansastraße Bahnhofsstraße	1.1
193	Du fährst gerade die Bahnhofstraße entlang	Du folgst der Bahnhofstraße	1.1
194	Das ist die T-Kreuzung zwischen der Bahnhofstraße und der Schulstraße	Das ist die T-Kreuzung Bahnhofstraße und Schulstraße	1.1
195	Du befindest dich gerade auf einer freien Fläche	Das ist ein Platz	?
196	Das ist der linke Kartenrand gewesen	Das ist der linke Kartenrand	1.1
197	Du bist wieder auf der freien Fläche	Das ist (wieder) der Platz	?
198	Die freie Fläche ist begrenzt durch die Schulstraße, der alten Gasse, Tannenstraße und der Silberstraße. Auf dieser Fläche befindet dich auch noch eine Schule, eine Gebäude-Landmarke	Der Platz ist begrenzt durch Schulstraße, Alte Gasse, Tannenstraße und Silberstraße.	4.1
199	”	Auf dem Platz, der durch Schulstraße, Alte Gasse, Tannenstraße und Silberstraße begrenzt ist, befindet sich auch die Schule	3.2
200	Jetzt bist die wieder am Kartenrand	Das ist der Kartenrand	1.1
201	Schulstraße	Das ist die Schulstraße	1.1
202	Und zwar am nördlichen Ende der Schulstraße	Das ist das nördliche Ende der Schulstraße	1.1

No.	Utterance	Split Utterance Analyzed	Message Class
203	Du hast gerade die T-Kreuzung zur Kleinen Gasse passiert	Das ist die T-Kreuzung Schulstraße Kleine Gasse	1.1
204	Eine weitere T-Kreuzung zur Alten Gasse	Das ist die T-Kreuzung Schulstraße Alte Gasse	1.1
205	Du bist immer noch in der Schulstraße	Das ist (immer noch) die Schulstraße	1.1
206	Das ist die Kleine Gasse	Das ist die Kleine Gasse	1.1
207	Die Kleine Gasse ist mit einer T-Kreuzung mit der Schulstraße verbunden	Die kleine Gasse ist in einer T-Kreuzung mit der Schulstraße verbunden	2.2
208	Die Kleine Gasse endet links in einer Sackgasse	Die Kleine Gasse endet links in einer Sackgasse	2.1
209	Das ist der obere Kartenrand	Das ist der obere Kartenrand	1.1
210	Das ist die Tannenstraße	Das ist die Tannenstraße	1.1
211	Das eben war die Ecke Tannenstraße, Schulstraße	Das ist die Ecke Tannenstraße, Schulstraße	1.1
212	Jetzt bist du wieder in der Schulstraße	Das ist die Schulstraße	1.1
213	Südlich von dir befindet sich eine T-Kreuzung, die die Schulstraße mit der Kleinen Gasse verbindet	Südlich von dir befindet sich die T-Kreuzung Schulstraße Kleine Gasse	1.2
214	Nun bist du wieder am oberen Kartenrand	Das ist der obere Kartenrand	1.1
215	Das ist die Ecke zwischen der Tannenstraße und der Schulstraße	Das ist die Ecke Tannenstraße, Schulstraße	1.1
216	Das ist immer noch die Ecke zwischen der Tannenstraße und der Schulstraße	Das ist die Ecke Tannenstraße, Schulstraße	1.1
217	Das ist die Tannenstraße	Das ist die Tannenstraße	1.1
218	Jetzt bist du in die Silberstraße abgebogen	Jetzt bist du in die Silberstraße abgebogen	1.1
219	Du bist wieder in der Schulstraße gelandet	Das ist die Schulstraße	1.1
220	Über dir befindet sich eine T-Kreuzung, die Schulstraße mit Alter Gasse verbindet	Die T-Kreuzung Schulstraße Alte Gasse ist über dir	1.2

No.	Utterance	Split Utterance Analyzed	Message Class
221	Jetzt bist du auf einer freien Fläche	Das ist ein Platz	1.1
222	Rechts von dir ist die Schulstraße	Die Schulstraße liegt rechts von dir	1.2
223	Jetzt bist du wieder in der Bahnhofstraße, die Bahnhofstraße ist die südlichste Straße und endet links in einer Sackgasse	Das ist die Bahnhofstraße	1.1
224	"	Die Bahnhofstraße ist die südlichste Straße	3.3
225	"	Die Bahnhofstraße endet links in einer Sackgasse	2.1
226	Jetzt bist du kurz vor der Ecke Bahnhofstraße, Hansastraße	Jetzt bist du kurz vor der Ecke Bahnhofstraße, Hansastraße	1.3
227	Das ist die Hansastraße	Das ist die Hansastraße	1.1
228	Das ist die T-Kreuzung zwischen Hansastraße und Tannenstraße	Das ist die T-Kreuzung Hansastraße Tannenstraße	1.1
229	Das ist die Sackgasse, in der die Tannenstraße rechts endet	Das ist die Sackgasse in der die Tannenstraße rechts endet	1.1
230	Südlich der Tannenstraße liegt noch die Landmarke Kirche	Südlich der Tannenstraße liegt noch die Kirche	1.3
231	Du bist wieder in der Bahnhofstraße	Das ist die Bahnhofstraße	1.1
232	Schulstraße	Das ist die Schulstraße	1.1
233	Wieder in der Tannenstraße	Das ist die Tannenstraße	1.1
234	Das ist die Ecke Schulstraße, Tannenstraße	Das ist die Ecke Schulstraße, Tannenstraße	1.1
235	Das war die T-Kreuzung zur Silberstraße	Das ist die Kreuzung Tannenstraße Silberstraße	1.1
236	Jetzt bist du am rechten Kartenrand	Das ist der rechte Kartenrand	1.1
237	Das ist die Hansastraße	Das ist die Hansastraße	1.1
238	Links, nördlich von dir befindet sich die Landmarke Kirche	Links, nördlich von dir befindet sich die Kirche	1.2

No.	Utterance	Split Utterance Analyzed	Message Class
239	Jetzt befindet sich die Kirche südlich von dir und du bist in der Tannenstraße	Jetzt befindet sich die Kirche südlich von dir	1.2
240	”	Das ist die Tannenstraße	1.1
241	Das ist die Silberstraße	Das ist die Silberstraße	1.1
242	Rechts der Silberstraße befinden sich drei Bäume	Rechts von der Silberstraße sind drei Bäume	1.2
243	Du befindest dich jetzt auf der Dorfstraße	Das ist die Dorfstraße	1.1
244	Jetzt bist du gerade in den Kirchweg eingebogen	Das ist der Kirchweg	1.1
245	Südlich von dir, äähh jetzt verlässt du die Karte. Südlich von dir befindet sich die Kirche	Südlich von dir befindet sich die Kirche	1.2
246	Das ist ääähhh, die obere linke Ecke der Karte	Das ist die obere linke Ecke der Karte	1.1
247	Jetzt bist du gerade an der Schule vorbeigegangen	Das war die Schule	1.1
248	Du befindest dich wieder auf der Dorfstraße	Das ist die Dorfstraße	1.1
249	Jetzt befindest du dich auf dem Kirchweg und kommst zur Kreuzung zur Dorfstraße	Kirchweg und Dorfstraße kreuzen sich	2.2
250	Jetzt bist du wieder auf der Dorfstraße angelangt	Das ist die Dorfstraße	1.1
251	Jetzt bist du die Dorfstraße entlanggefahren am Dorfplatz vorbei	Die Dorfstraße begrenzt den Dorfplatz	4.1
252	”	Das ist die Dorfstraße	1.1
253	Und kannst, bist jetzt rechts in den Kirchweg eingebogen	Das ist der Kirchweg	1.1
254	”	Der Kirchweg ist mit der Dorfstraße verbunden	2.2
255	Du kannst in äähh, hier kannst du südlich in die Dorfstraße und dann links in die Marktstraße	Du befindest dich an der Kreuzung Dorfstraße, Marktstraße	1.1
256	Wo du jetzt südlich in die alte Gasse abbiegen kannst	Das ist die Marktstraße	1.1

No.	Utterance	Split Utterance Analyzed	Message Class
257	”	Das ist die Kreuzung alte Gasse Markststraße	1.1
258	Dann direkt T-Kreuzung südlich der alten Gasse kommst du zur neuen Gasse	Alte Gasse und neue Gasse besitzen eine gemeinsame T-Kreuzung	2.2
259	Südlich der neuen Gasse ist die Schule	Südlich der neuen Gasse ist die Schule	1.2
260	Und die Hochstraße, das befindet sich südwestlich der neuen Gasse	Die Hochstraße ist südwestlich der neuen Gasse	1.2
261	Das ist die Straße	Das ist die Hochstraße	1.1
262	Jetzt befindest du dich auf der Schule	Das ist die Schule	1.1
263	Jetzt bist du wieder auf der neuen Gasse	Das ist die neue Gasse	1.1
264	Hochstraße	Das ist die Hochstraße	1.1
265	Alte Gasse	Das ist die alte Gasse	1.1
266	Der Marktstraße	Das ist die Marktstraße	1.1
267	Wieder auf der Dorfstraße	Das ist die Dorfstraße	1.1
268	Jetzt bist du wieder auf dem Kirchweg	Das ist der Kirchweg	1.1
269	Südlich befindet sich die Kirche	Südlich des Kirchweges ist die Kirche	1.2
270	Südlich davon der Dorfplatz	Südlich des Kirchweges ist der Dorfplatz	1.2
271	Jetzt bist du auf der Kirche	Das ist die Kirche	1.1
272	Südlich davon befindet sich der Dorfplatz	Südlich davon Befindet sich der Dorfplatz	1.2
273	Und wieder die Dorfstraße	Das ist die Dorfstraße	1.1
274	Das ist die Kirche	Das ist die Kirche	1.1
275	Die Dorfstraße	Das ist die Dorfstraße	1.1
276	Jetzt befindest du dich wieder auf der Hochstraße	Das ist die Hochstraße	1.1
277	Neue Gasse	Das ist die neue Gasse	1.1
278	Südlich ist die Schule	Südlich von dir/von der neuen Gasse ist die Schule	1.2
279	Du explorierst den nördlichen Kartenrand	Das ist der nördliche Kartenrand	1.1

No.	Utterance	Split Utterance Analyzed	Message Class
280	Du explorierst den rechten und den südlichen Kartenrand, und den linken Kartenrand	Das ist der östlichen Kartenrand	1.1
281	"	Das ist der südliche Kartenrand	1.1
282	"	Das ist der westliche Kartenrand	1.1
283	Das ist die Hansastraße	Das ist die Hansastraße	1.1
284	Das ist die Mühlenstraße	Das ist die Mühlenstraße	1.1
285	Die Mühlenstraße ist ganz nördlich in der Karte	Die Mühlenstraße ist die nördlichste Straße auf der Karte	3.3
286	Das ist die Viktoriastraße, die ist ganz rechts in der Karte	Das ist die Viktoriastraße	1.1
287	"	Die Viktoria Straße ist ganz rechts auf der Karte	3.3
288	Das ist die Feldstraße, die ist ganz südlich in der Karte und mündet links in einer Sackgasse die du gerade explorierst	Das ist die Feldstraße	1.1
289	"	Die Feldstraße ist die südlichste der Straßen	3.3
290	"	Die Feldstraße endet links in einer Sackgasse	2.1
291	"	Das ist die Sackgasse, in der die Feldstraße endet	1.1
292	Das ist wieder die Hansastraße ganz links in der Karte, parallel zum Rand	Das ist die Hansastraße	1.1
293	"	Die Hansastraße ist ganz links auf der Karte	1.1
294	"	Die Hansastraße parallel zum Kartenrand	2.3
295	Jetzt bist du auf den Leinpfad	Das ist der Leinpfad	1.1
296	Jetzt bist du auf der Lindenstraße	Das ist die Lindenstraße	1.1
297	Die Lindenstraße ist parallel zur Mühlenstraße	Die Lindenstraße ist parallel zur Mühlenstraße	2.3

No.	Utterance	Split Utterance Analyzed	Message Class
298	Jetzt bist du auf der Viktoriastraße	Das ist die Viktoriastraße	1.1
299	Jetzt bist du wieder auf der Mühlenstraße, die rechts in einer Sackgasse endet, die du gerade explorierst	Das ist die Mühlenstraße	1.1
300	”	Die Mühlenstraße endet rechts in einer Sackgasse	2.1
301	”	Das ist die Sackgasse, in der die Mühlenstraße endet	1.1
302	Jetzt bist du wieder auf der Viktoriastraße	Das ist die Viktoriastraße	1.1
303	Jetzt bist du wieder auf der Lindenstraße	Das ist die Lindenstraße	1.1
304	Jetzt bist du wieder auf der Feldstraße, die südlich in der Karte liegt und parallel zum unteren Kartenrand verläuft, gerade warst du kurz in der neuen Gasse, die Von der Viktoriastraße die du gerade explorierst abgeht.	Das ist die Feldstraße	1.1
305	”	Die Feldstraße liegt südlich in der Karte	3.1
306	”	Die Feldstraße verläuft parallel zum Kartenrand	2.3
307	”	Das ist die neue Gasse	1.1
308	”	Die neue Gasse geht von der Viktoriastraße ab	2.2
309	”	Das ist die Viktoriastraße	1.1
310	Du bist wieder auf der Feldstraße	Das ist die Feldstraße	1.1
311	Du bist wieder auf der Viktoriastraße	Das ist die Viktoriastraße	1.1
312	Du bist wieder auf der Hansastraße	Das ist die Hansastraße	1.1
313	Du bist wieder auf der Mühlenstraße	Das ist die Mühlenstraße	1.1
314	Du bist wieder auf der Lindenstraße	Das ist die Lindenstraße	1.1

No.	Utterance	Split Utterance Analyzed	Message Class
315	Du bist wieder auf der Viktoriastraße	Das ist die Viktoriastraße	1.1
316	Du bist wieder auf der Feldstraße	Das ist die Feldstraße	1.1
317	Du bist wieder auf der Lindenstraße	Das ist die Lindenstraße	1.1
318	Du bist wieder auf der Viktoriastraße	Das ist die Viktoriastraße	1.1
319	Du bist wieder auf der Feldstraße	Das ist die Feldstraße	1.1
320	Du bist wieder auf der Hansastraße	Das ist die Hansastraße	1.1
321	Jetzt bist du auf der Mühlenstraße	Das ist die Mühlenstraße	1.1
322	Jetzt bist du auf dem Leinpfad	Das ist der Leinpfad	1.1
323	Jetzt explorierst du den Bahnhofsplatz, jetzt explorierst du den Bahnhof	Das ist der Bahnhof	1.1
324	Auf dem Bahnhofsplatz liegt noch ein weiteres Gebäude und drei Bäume	Auf dem Bahnhofsplatz ist ein weiteres Gebäude	3.2
325	”	Auf dem Bahnhofsplatz sind drei Bäume	3.2
326	Jetzt explorierst du die Hansastraße	Das ist die Hansastraße	1.1
327	Jetzt bist du wieder, jetzt explorierst du den Bahnhof wieder	Das ist der Bahnhof	1.1
328	Das ist der Bahnhof	Das ist der Bahnhof	1.1
329	Das ist eine freie Fläche auf der keine weiteren Landm äääähh, keine weiteren Gebäude oder Bäume oder sonst irgendwelche Kartenteile liegen	Das ist eine freie Fläche ohne weitere Gebäude oder Bäume	?
330	Du bist wieder auf der Hansastraße	Das ist die Hansastraße	1.1
331	Du bist auf der Lindenstraße	Das ist die Lindenstraße	1.1
332	Du bist auf der Feldstraße	Das ist die Feldstraße	1.1

No.	Utterance	Split Utterance Analyzed	Message Class
333	Über dir liegen drei Bäume	Über dir liegen drei Bäume	1.2
334	Jetzt sind die Bäume links neben dir	Jetzt sind die Bäume links neben dir	1.2
335	Das war ein Baum	Das war ein Baum	1.1
336	Das ist der Bahnhofsplatz	Das ist der Bahnhofsplatz	1.1
337	Das ist die Lindenstraße	Das ist die Lindenstraße	1.1
338	Das ist die Feldstraße	Das ist die Feldstraße	1.1
339	Rechts von dir sind die Bäume	Rechts von dir sind die Bäume	1.2
340	Du bist im unteren der Bäume, du bist im mittleren der Bäume	Das ist der untere der Bäume	1.1
341	”	Das ist der mittleren der Bäume	1.1
342	Jetzt bist du wieder in der Feldstraße	Das ist die Feldstraße	1.1
343	Das ist der Mittlere der Bäume	Das ist der mittlere der Bäume	1.1
344	Das ist die Viktoriastraße	Das ist die Viktoriastraße	1.1
345	Das ist der Mittlere der Bäume	Das ist der mittleren der Bäume	1.1
346	Das ist der Bahnhofsplatz	Das ist der Bahnhofsplatz	1.1
347	Das ist die Lindenstraße	Das ist die Lindenstraße	1.1
348	Das ist die Viktoriastraße, das ist die Mühlenstraße. Am linken Ende liegt südlich dieser Straße noch eine Landmarke an der Ecke zur Hansastraße, die jetzt unter dir ist	Das ist die Viktoriastraße	1.1
349	”	Das ist die Mühlenstraße	1.1
350	”	Die Viktoriastraße ist zu ihrer linken mit der Hansastraße über Eck verbunden	2.2
351	”	Südlich dieser Verbindung liegt eine weitere Landmarke	1.2
352	Die Landmarke liegt südlich, jetzt liegt sie rechts von dir, rechts über dir	Die Landmarke liegt jetzt unter dir	5

No.	Utterance	Split Utterance Analyzed	Message Class
353	”	Die Landmarke liegt jetzt südlich von dir	5
354	”	Die Landmarke liegt jetzt rechts von dir	5
355	”	Die Landmarke liegt jetzt rechts über von dir	5
356	Jetzt bist du auf dem Bahnhofsplatz	Das ist der Bahnhofsplatz	1.1
357	Jetzt bist du auf der Lindenstraße	Das ist die Lindenstraße	1.1
358	Jetzt bist du auf dem Leinpfad	Das ist der Leinpfad	1.1
359	Und auf der Mühlenstraße	Das ist die Mühlenstraße	1.1
360	Jetzt bist du wieder auf dem Bahnhof	Das ist der Bahnhof	1.1
361	Links an von der Mühlenstraße, es gibt noch eine Landm noch ein Gebäude in der Karte was du noch nicht exploriert hast, das liegt jetzt links, genau links neben dir	Es gibt noch ein Gebäude, auf der Karte das du noch nicht exploriert hast	5
362	”	Das Gebäude liegt links von der Mühlenstraße	1.2
363	Da du warst fast dran, rechts von dir, etwas etwas etwas nördlicher	Das Gebäude liegt jetzt rechts neben dir	5
364	”	Das Gebäude liegt jetzt nördlich von dir	5
365	Etwas, so jetzt bist du wieder auf der Mühlenstraße, jetzt gerade nach unten, ja jetzt bist du leider knapp dran vorbei, etwas nach links etwas nach oben	Das ist die Mühlenstraße	1.1
366	”	Das Gebäude liegt jetzt südlich von dir	5
367	”	Das Gebäude liegt jetzt links neben dir	5
368	Jetzt explorierst du die Schule	Das ist die Schule	1.1

No.	Utterance	Split Utterance Analyzed	Message Class
369	Du hast jetzt alles was auf der Karte dargestellt wird exploriert		1.1
370	Du explorierst gerade den Kartenrand	Das ist der Kartenrand	1.1
371	Jetzt bist du in der neuen Gasse, die eine Sackgasse ist	Das ist die neue Gasse	1.1
372	"	Die neue Gasse ist eine Sackgasse	2.1
373	kommst auf einer T-Kreuzung	Die neue Gasse hat eine T-Kreuzung mit der Viktoriastraße	2.2
374	Du warst auf der Viktoriastraße	Das ist die Viktoriastraße	1.1
375	Die Viktoriastraße verläuft von Norden nach Süden	Die Viktoriastraße verläuft von Norden nach Süden	2.3
376	T-Kreuzung	Das ist die T-Kreuzung von Viktoriastraße und Lindenstraße	1.1
377	Lindenstraße	Das ist die Lindenstraße	1.1
378	Wieder auf Viktoriastraße	Das ist die Viktoriastraße	1.1
379	T-Kreuzung Mühlenstraße	Das ist die T-Kreuzung Viktoriastraße Mühlenstraße	1.1
380	Die Mühlenstraße verläuft Ost nach West	Die Mühlenstraße verläuft Ost nach West	2.3
381	Endet im Osten in einer Sackgasse und ist die nördlichste der Straßen	Die Mühlenstraße endet rechts in einer Sackgasse	2.1
382	"	Die Mühlenstraße ist die nördlichste Straße	3.3
383	Jetzt bist du an der Ecke Mühlenstraße Hansastraße	Die Mühlenstraße ist mit der Hansastraße über eine Ecke verbunden	2.2
384	Die Hansastraße verläuft von Norden nach Süden und ist die am weitesten links gelegene Straße	Die Hansastraße verläuft von Norden nach Süden	2.3

No.	Utterance	Split Utterance Analyzed	Message Class
385	”	Die Hansastraße ist die am weitesten links gelegene Straße, die von Nord nach Süd verläuft	3.3
386	T-Kreuzung	Das ist die T-Kreuzung verbunden	1.1
387	Das ist eine Sackgasse, die zur Feldstraße gehört	Das ist die Sackgasse, in der die Feldstraße endet	1.1
388	'Die Feldstraße verläuft von links nach rechts und ist die am weitesten im Süden gelegene Straße	Die Feldstraße verläuft von links nach rechts	2.3
389	”	Die Feldstraße ist die südlichste Straße	3.3

Appendix B

Empirical Evaluation: Materials

B.1 Assisting Utterances given in Experiments 1 and 2

The following assisting utterances are the full set of prerecorded assisting utterances used in Experiments 1 and 2. In Experiment 3, only the assisting utterances related to the track network were used. The corresponding utterances for name set 2 matched exactly, aside from the names used.

Map 1 with name set 1

Kartenrand
Das ist der obere Kartenrand
Das ist der linke Kartenrand
Das ist der untere Kartenrand
Das ist der rechte Kartenrand

Tracks/Straßen

Poststraße
Das ist die Poststraße
Die Poststraße ist von Kartenrand oben und links begrenzt
Von der Postraße geht rechts die Humboldtstraße ab

Poststraße oberer Teil
Rechts neben diesem Teil der Poststraße steht eine Eiche

Poststraße linker Teil
Unter diesem Teil der Poststraße liegt die Bertholt-Brecht-Schule

Humboldtstraße
Das ist die Humboldtstraße
Die Humboldtstraße endet links in einer T-Kreuzung mit der Poststraße und wird rechts durch den Kartenrand begrenzt

Die Humboldtstraße hat zwei Kreuzungen: Mit der Goethestraße und mit der Hegelstraße

Humboldtstraße, rechter Teil (nach der ersten Kreuzung von links)
Dieser Teil der Humboldtstraße ist parallel zum Lärchenweg
Über diesem Teil der Humboldtstraße liegen die Christuskirche und die Universität

Goethestraße
Das ist die Goethestraße
Die Goethestraße endet oben in einer Ecke mit dem Lärchenweg und wird unten durch den Kartenrand begrenzt
Die Goethestraße ist parallel zur Hegelstraße
Die Goethestraße kreuzt die Humboldtstraße und ist durch eine Ecke mit dem Lärchenweg verbunden

Hegelstraße
Das ist die Hegelstraße
Die Hegelstraße ist parallel zur Goethestraße
Die Hegelstraße kreuzt die Humboldtstraße und den Lärchenweg
Die Hegelstraße ist oben und unten durch den Kartenrand begrenzt
An der Hegelstraße liegen vier Gebäude

Hegelstraße zw. Christuskirche und Universität
Jetzt bist du zwischen Christuskirche und Universität
Hegelstraße zwischen Aldi und Hauptbahnhof
Jetzt bist du zwischen Aldi und Hauptbahnhof

Lärchenweg
Das ist der Lärchenweg
Der Lärchenweg ist parallel zum rechten Teil der Humboldtstraße
Der Lärchenweg endet links in einer Ecke mit der Goethestraße und rechts in einer Sackgasse
Der Lärchenweg kreuzt die Hegelstraße und ist durch eine Ecke mit der Goethestraße verbunden
Über dem Lärchenweg liegen der Aldi und der Hauptbahnhof
Lärchenweg Sackgasse rechts
Das ist die Sackgasse, in der der Lärchenweg rechts endet

Kreuzungen, Einmündungen, Ecken

Einmündung Poststraße Humboldtstraße
Das ist die Einmündung der Humboldtstraße in die Poststraße

Kreuzung Humboldtstraße Goethestraße
Das ist die Kreuzung Humboldtstraße Goethestraße

Kreuzung Humboldtstraße Hegelstraße
Das ist die Kreuzung Humboldtstraße Hegelstraße
Links oberhalb der Kreuzung Humboldtstraße Hegelstraße liegt die Christuskirche und rechts oberhalb die Universität

Kreuzung Hegelstraße Lärchenweg
Das ist die Kreuzung Hegelstraße Lärchenweg
Links oberhalb der Kreuzung Hegelstraße Lärchenweg liegt der Aldi und rechts oberhalb liegt der Hauptbahnhof

Ecke Goethestraße Lärchenweg
In dieser Ecke treffen sich Goethestraße und Lärchenweg

Landmarken

Eiche
Das ist die Eiche
Die Eiche liegt rechts von der Poststraße
Die Eiche liegt im oberen Bereich der Karte

Bertholt-Brecht-Schule
Das ist die Bertholt-Brecht-Schule
Die Bertholt-Brecht-Schule liegt unterhalb der Poststraße
Die Bertholt-Brecht-Schule liegt im unteren Bereich der Karte

Christuskirche
Das ist die Christuskirche
Rechts unterhalb der Christuskirche liegt die Kreuzung Humboldtstraße Hegelstraße
Rechts von der Christuskirche liegt die Universität

Universität
Das ist die Universität
Links unterhalb der Universität liegt die Kreuzung Hegelstraße Humboldtstraße
Links von der Universität liegt die Christuskirche

Aldi
Das ist der Aldi
Rechts unterhalb des Aldis liegt die Kreuzung Lärchenweg Hegelstraße
Rechts vom Aldi liegt der Hauptbahnhof

Hauptbahnhof
Das ist der Hauptbahnhof
Links unterhalb des Hauptbahnhofes liegt die Kreuzung Lärchenweg Hegelstraße
Links vom Hauptbahnhof liegt der Aldi

Map 2 with name set 1

Kartenrand

Das ist der obere Kartenrand
Das ist der linke Kartenrand
Das ist der untere Kartenrand
Das ist der rechte Kartenrand

Tracks/Straßen

Poststraße
Das ist die Poststraße
Die Poststraße ist links vom Kartenrand begrenzt und mündet rechts in die Hegelstraße ein

Poststraße linker Teil
Rechts oberhalb dieses Teils der Poststraße liegt die Universität

Poststraße mittlerer Teil
Links unter diesem Teil der Poststraße liegt die Eiche

Hegelstraße
Das ist die Hegelstraße
Die Hegelstraße ist oben und unten durch den Kartenrand begrenzt
Die Hegelstraße hat zwei Kreuzungen, mit dem Lärchenweg und der Humboldtstraße, und eine Einmündung von der Poststraße
Die Hegelstraße ist parallel zur Goethestraße
An der Hegelstraße liegen vier Gebäude

Hegelstraße, zwischen Bertholt-Brecht-Schule und Hauptbahnhof
Jetzt bist du zwischen Bertholt-Brecht-Schule und Hauptbahnhof
Hegelstraße, zwischen Aldi und Christuskirche
Jetzt bist du zwischen Aldi und Christuskirche

Goethestraße
Das ist die Goethestraße

Die Goethestraße ist parallel zur Hegelstraße
Die Goethestraße wird oben durch den Kartenrand begrenzt und endet unten in einer Ecke mit dem Lärchenweg
In die Goethestraße mündet die Humboldtstraße ein

Humboldtstraße
Das ist die Humboldtstraße
Die Humboldtstraße ist parallel zur Lärchenweg
Die Humboldtstraße mündet links in die Goethestraße ein und endet rechts in einer Sackgasse
Die Humboldtstraße kreuzt die Hegelstraße
Über der Humboldtstraße liegen der Aldi und die Christuskirche

Humboldtstraße, Sackgasse rechts
Das ist die Sackgasse, in der die Humboldtstraße rechts endet

Lärchenweg
Das ist der Lärchenweg
Der Lärchenweg ist parallel zur Humboldtstraße
Der Lärchenweg endet links in einer Ecke mit der Goethestraße und wird rechts durch den Kartenrand begrenzt
Der Lärchenweg kreuzt die Hegelstraße
Unter dem Lärchenweg liegen der Hauptbahnhof und die Bertholt-Brecht-Schule

Kreuzungen, Einmündungen, Ecken

Einmündung Poststraße Hegelstraße
Das ist die Einmündung der Poststraße in die Hegelstraße

Einmündung Humboldtstraße Goethestraße
Das ist die Einmündung der Humboldtstraße in die Goethestraße

Ecke Goethestraße Lärchenweg
Das ist die Ecke zwischen Goethestraße und Lärchenweg

Kreuzung Hegelstraße Lärchenweg
Das ist die Kreuzung zwischen Hegelstraße und Lärchenweg
Links unterhalb der Kreuzung Hegelstraße Lärchenweg liegt der Hauptbahnhof und rechts unterhalb die Bertholt-Brecht-Schule

Kreuzung Hegelstraße Humboldtstraße
Das ist die Kreuzung zwischen Hegelstraße und Humboldtstraße

Links oberhalb der Kreuzung Hegelstraße und Humboldtstraße liegt der Aldi und rechts oberhalb die Christuskirche

Landmarken

Universität
Das ist die Universität
Die Universität liegt rechts oberhalb der Poststraße
Die Universität liegt im oberen Bereich der Karte

Eiche
Das ist die Eiche
Die Eiche liegt links unterhalb der Poststraße
Die Eiche liegt in der unteren Hälfte der Karte

Hauptbahnhof
Das ist der Hauptbahnhof
Rechts oberhalb des Hauptbahnhofes liegt die Kreuzung Hegelstraße Lärchenweg
Rechts vom Hauptbahnhof liegt die Bertholt-Brecht-Schule

Bertholt-Brecht-Schule
Das ist die Bertholt-Brecht-Schule
Links oberhalb der Bertholt-Brecht-Schule liegt die Kreuzung Hegelstraße Lärchenweg
Links von der Bertholt-Brecht-Schule liegt der Hauptbahnhof

Aldi
Das ist der Aldi
Rechts unterhalb des Aldis liegt die Kreuzung Hegelstraße Humboldtstraße
Rechts vom Aldi liegt die Christuskirche

Christuskirche
Das ist die Christuskirche
Links unterhalb der Christuskirche liegt die Kreuzung Hegelstraße Humboldtstraße
Links von der Christuskirche liegt der Aldi

B.2 Instruction for Experiment 1

In den zwei folgenden Versuchen wird Ihnen jeweils eine andere haptische Karte mit dem PHANToM präsentiert. Sie haben 8 Minuten Zeit, diese zu erfühlen. Ihnen wird ein Hinweis gegeben, wenn die Zeit vorbei ist. Bitte merken Sie sich die Karten möglichst vollständig, inklusive der Namen der Straßen und anderen Kartenobjekte. Die Karten sind vergleichbar mit der unten abgebildeten, aber die Kartenobjekte haben "richtige" Namen. Bitte versuchen Sie, die Karte so zu lernen, dass Sie hinterher in der Lage sind, von jedem beliebigen Gebäude zu jedem anderen zu finden.

Im Anschluss an das Experiment werden Sie zu der Karte befragt. Es gibt die Antwortmöglichkeiten "Ja", "Nein" und "Ich weiß nicht". Zusätzlich werden Sie gebeten, Teile der Karte aus einem Set gegebener Teile, die teilweise fehlerhaft sind, zu identifizieren. Zwischen beiden Versuchen machen wir eine kurze Pause (ca. 10 Min.).

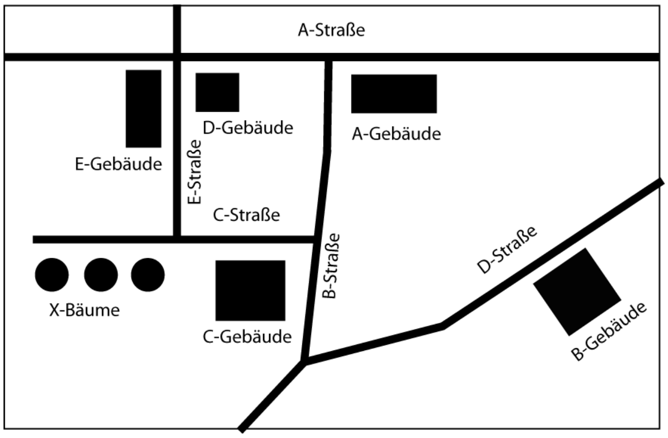


Figure B.1: The instruction used in Experiment 1.

B.3 Instruction for Experiment 3

In den folgenden drei Versuchsbedingungen wird Ihnen jeweils eine andere haptische Karte mit der haptischen Schnittstelle präsentiert. Sie haben 4 Minuten Zeit, diese zu erfühlen. Ihnen wird ein Hinweis gegeben, wenn die Zeit vorbei ist. Auf der Karte ist ein Straßennetz, es gibt keine weiteren Objekte. Bitte versuchen Sie, die Karte so zu lernen, dass sie das Straßennetz möglichst vollständig kennen.

Im Anschluss an das Experiment werden Sie zu dem Straßennetz befragt. Zu den Fragen gibt es die Antwortmöglichkeiten "Ja", "Nein" und "Ich weiß nicht". Weiter werden Sie gebeten, die Karte zu zeichnen. Wenn Ihnen Namensinformationen gegeben werden, bitten wir Sie, die Namen an die Straßen zu schreiben. Im Anschluss werden Sie gebeten, korrekte Teile der Karte unter falschen Teilen zu identifizieren.

Zwischen den Versuchsbedingungen machen wir jeweils eine kurze Pause.

Figure B.2: The instruction used in Experiment 3.

B.4 Configuration Questions: Experiments 1 and 2

Map 1, Name Set 1

Table B.1: The questions used for Map 1 with name set 1.

Question	Question (MAP1, SET1)	Correct Answer
1	Ist die Eiche auf der Karte links vom Hauptbahnhof?	t
2	Ist die Eiche auf der Karte links von der Hegelstraße?	t
3	Ist die Bertholt-Brecht-Schule auf der Karte links von der Christuskirche?	t
4	Ist der Hauptbahnhof auf der Karte über der Universität?	t
5	Ist die Christuskirche unterm Aldi?	t
6	Mündet die Humboldtstraße in die Poststraße ein?	t
7	Sind Goethestraße und Hegelstraße parallel?	t
8	Endet der Lärchenweg rechts in einer Sackgasse?	t
9	Kreuzen sich Hegelstraße und Lärchenweg?	t
10	Grenzen Goethestraße und Lärchenweg aneinander an?	t
11	Ist die Eiche rechts von der Christuskirche?	f
12	Ist die Hegelstraße links von der Bertholt-Brecht-Schule?	f
13	Ist die Bertholt-Brecht-Schule über der Eiche?	f
14	Ist der Hauptbahnhof unter dem Aldi?	f
15	Ist der Hauptbahnhof rechts von der Universität?	f
16	Mündet die Poststraße in die Hegelstraße ein?	f
17	Sind Humboldtstraße und Goethestraße parallel?	f
18	Endet die Goethestraße in einer Sackgasse?	f
19	Kreuzen sich Poststraße und Lärchenweg?	f
20	Grenzen Humboldstraße und Lärchenweg aneinander an?	f

Map 1, Name Set 2

Table B.2: The questions used for Map 1 with name set 2.

Question	Question (MAP1, SET2)	Correct Answer
1	Ist die Buche auf der Karte links von der Anne-Frank-Schule?	t
2	Ist die Buche auf der Karte links von der Blumenstraße?	t
3	Ist das Museum links von der Gedächtniskirche?	t
4	Ist die Anne-Frank-Schule über dem Rathaus?	t
5	Ist die Gedächniskirche unterm Lidl?	t
6	Mündet die Hochstraße in die Bergstraße ein?	t
7	Sind Amselweg und Blumenstraße parallel?	t
8	Endet die Dorfstraße rechts in einer Sackgasse?	t
9	Kreuzen sich Blumenstraße und Dorfstraße?	t
10	Grenzen Amselweg und Dorfstraße aneinander an?	t
11	Ist die Buche rechts von der Gedächtniskirche?	f
12	Ist die Blumenstraße links vom Museum?	f
13	Ist das Museum über der Buche?	f
14	Ist die Anne-Frank-Schule unter dem Lidl?	f
15	Ist die Anne-Frank-Schule rechts vom Rathaus?	f
16	Mündet die Bergstraße in die Blumenstraße ein?	f
17	Sind Hochstrraße und Amselweg parallel?	f
18	Endet der Amselweg in einer Sackgasse?	f
19	Kreuzen sich Bergstraße und Dorfstraße?	f
20	Grenzen Hochstraße und Dorfstraße aneinander an?	f

Map 2, Name Set 1

Table B.3: The questions used for Map 2 with name set 1. (* This is the formulation used in Experiment 2. Question five was excluded in the analysis of Experiment 1. The formulation in Experiment 1 was: 'Ist die Christuskirche in der Karte unter dem Hauptbahnhof?')

Question	**Question (MAP2, SET1)**	**Correct Answer**
1	Ist die Universität auf der Karte links von der Christuskirche?	t
2	Ist die Universität auf der Karte links von der Hegelstraße?	t
3	Ist der Aldi über dem Hauptbahnhof?	t
4	Ist die Christuskirche über der Eiche?	t
5	Ist die Christuskirche über dem Hauptbahnhof?*	t
6	Mündet die Poststraße in die Hegelstraße ein?	t
7	Sind Humboldtstraße und Lärchenweg parallel?	t
8	Endet die Humboldtstraße rechts in einer Sackgasse?	t
9	Kreuzen sich Humboldtstraße und Hegelstraße?	t
10	Grenzen Humboldtstraße und Goethestraße direkt aneinander an?	t
11	Ist die Universität auf der Karte rechts vom Aldi?	f
12	Ist die Poststraße über der Universität?	f
13	Ist der Aldi unter der Bertholt-Brecht-Schule?	f
14	Ist die Christuskirche auf der Karte unter der Eiche?	f
15	Ist die Bertholt-Brecht-Schule links von der Eiche?	f
16	Mündet die Hegelstraße in die Goethestraße ein?	f
17	Sind Lärchenweg und Goethestraße parallel?	f
18	Endet der Lärchenweg in einer Sackgasse?	f
19	Kreuzen sich Goethestraße und Poststraße?	f
20	Grenzen Humboldtstraße und Lärchenweg aneinander an?	f

Map 2, Name Set 2

Table B.4: The questions used for Map 2 with name set 2. (* This is the formulation used in Experiment 2. Question five was excluded in the analysis of Experiment 1. The formulation in Experiment 1 was: 'Ist die Gedächtniskirche in der Karte über der Anne-Frank-Schule?')

Question	**Question (MAP2, SET2)**	**Correct Answer**
1	Ist die Gedächtniskirche auf der Karte links von der Anne-Frank-Schule?	t
2	Ist die Gedächtniskirche auf der Karte links von der Bergstraße?	t
3	Ist der Lidl über dem Rathaus?	t
4	Ist die Anne-Frank-Schule in der Karte über der Buche?	t
5	Ist die Anne-Frank-Schule in der Karte über dem Rathaus?*	t
6	Mündet die Blumenstraße in die Bergstraße ein?	t
7	Sind Hochstraße und Dorfstraße parallel?	t
8	Endet die Hochstraße rechts in einer Sackgasse?	t
9	Kreuzen sich Hochstraße und Bergstraße?	t
10	Grenzen Hochstraße und Amselweg direkt aneinander an?	t
11	Ist die Gedachtniskirche auf der Karte rechts vom Lidl?	f
12	Ist die Blumenstraße über der Gedächtniskirche?	f
13	Ist der Lidl unter dem Museum?	f
14	Ist die Anne-Frank-Schule auf der Karte unter der Buche?	f
15	Ist das Museum links von der Buche?	f
16	Mündet die Bergstraße in den Amselweg ein?	f
17	Sind Dorfstraße und Amselweg parallel?	f
18	Endet die Dorfstraße in einer Sackgasse?	f
19	Kreuzen sich Amselweg und Blumenstraße?	f
20	Grenzen Hochstraße und Dorfstraße aneinander an?	f

B.5 Configuration Questions: Experiment 3

Map 1, Name Set 1

Table B.5: The questions used for Map 1 with name set 1.

Question	Question (MAP1, SET1)	Correct Answer
1	Kreuzen sich Goethestraße und Humboldtstraße?	t
2	Sind Lärchenweg und Humboldtstraße teilweise oder ganz parallel?	t
3	Wird die Hegelstraße oben durch den Kartenrand begrenzt?	t
4	Kreuzen sich Humboldtstraße und Hegelstraße?	t
5	Wird die Humboldtstraße rechts durch den Kartenrand begrenzt?	t
6	Mündet die Humboldtstraße in die Poststraße ein?	t
7	Sind Goethestraße und Hegelstraße teilweise oder ganz parallel?	t
8	Endet der Lärchenweg rechts in einer Sackgasse, wird also nicht durch den Kartenrand begrenzt?	t
9	Kreuzen sich Hegelstraße und Lärchenweg?	t
10	Berühren sich Goethestraße und Lärchenweg direkt?	t
11	Kreuzen sich Goethestraße und Hegelstraße?	f
12	Sind Lärchenweg und Hegelstraße teilweise oder ganz parallel?	f
13	Wird die Humboldtstraße links durch den Kartenrand begrenzt?	f
14	Kreuzen sich Lärchenweg und Humboldtstraße?	f
15	Wird der Lärchenweg rechts durch den Kartenrand begrenzt?	f
16	Mündet die Poststraße in die Hegelstraße ein?	f
17	Sind Humboldtstraße und Goethestraße teilweise oder ganz parallel?	f
18	Endet die Goethestraße oben in einer Sackgasse, wird also nicht durch den Kartenrand begrenzt?	f
19	Kreuzen sich Poststraße und Lärchenweg?	f
20	Berühren sich Humboldtstraße und Lärchenweg direkt?	f

Map 1, Name Set 2

Table B.6: The questions used for Map 1 with name set 2.

Question	Question (MAP1, SET2)	Correct Answer
1	Kreuzen sich Amselweg und Hochstraße?	t
2	Sind Dorfstraße und Hochstraße teilweise parallel?	t
3	Wird die Blumenstraße oben durch den Kartenrand begrenzt?	t
4	Kreuzen sich Hochstraße und Blumenstraße?	t
5	Wird die Hochstraße rechts durch den Kartenrand begrenzt?	t
6	Mündet die Hochstraße in die Bergstraße ein?	t
7	Sind Amselweg und Blumenstraße teilweise oder ganz parallel?	t
8	Endet die Dorfstraße rechts in einer Sackgasse, wird also nicht durch den Kartenrand begrenzt?	t
9	Kreuzen sich Blumenstraße und Dorfstraße?	t
10	Berühren sich Amselweg und Dorfstraße direkt?	t
11	Kreuzen sich Amselweg und Blumenstraße?	f
12	Sind Dorfstraße und Blumenstraße teilweise oder ganz parallel?	f
13	Wird die Hochstraße links durch den Kartenrand begrenzt?	f
14	Kreuzen sich Dorfstraße und Hochstraße?	f
15	Wird die Dorfstraße rechts durch den Kartenrand begrenzt?	f
16	Mündet die Bergstraße in die Blumenstraße ein?	f
17	Sind Hochstraße und Amselweg parallel?	f
18	Endet der Amselweg oben in einer Sackgasse, wird also nicht durch den Kartenrand begrenzt?	f
19	Kreuzen sich Bergstraße und Dorfstraße?	f
20	Berühren sich Hochstraße und Dorfstraße direkt?	f

Map 2, Name Set 1

Table B.7: The questions used for Map 2 with name set 1.

Question	Question (MAP2, SET1)	Correct Answer
1	Kreuzen sich Lärchenweg und Hegelstraße?	t
2	Sind Goethestraße und Hegelstraße teilweise oder ganz parallel?	t
3	Wird die Hegelstraße oben durch den Kartenrand begrenzt?	t
4	Treffen sich Goethestraße und Lärchenweg in einer Ecke?	t
5	Wird der Lärchenweg rechts durch den Kartenrand begrenzt?	t
6	Mündet die Poststraße in die Hegelstraße ein?	t
7	Sind Humboldtstraße und Lärchenweg teilweise oder ganz parallel?	t
8	Endet die Humboldtstraße rechts in einer Sackgasse, wird also nicht durch den Kartenrand begrenzt?	t
9	Kreuzen sich Humboldtstraße und Hegelstraße?	t
10	Berühren sich Humboldtstraße und Goethestraße direkt?	t
11	Kreuzen sich Goethestraße und Hegelstraße?	f
12	Sind Humboldtstraße und Goethestraße teilweise oder ganz parallel?	f
13	Wird die Poststraße rechts durch den Kartenrand begrenzt?	f
14	Kreuzen sich Lärchenweg und Humboldtstraße?	f
15	Wird die Humboldtstraße rechts durch den Kartenrand begrenzt?	f
16	Mündet die Hegelstraße in die Goethestraße ein?	f
17	Sind Lärchenweg und Goethestraße parallel?	f
18	Endet der Lärchenweg in einer Sackgasse, wird also nicht durch den Kartenrand begrenzt?	f
19	Kreuzen sich Goethestraße und Poststraße?	f
20	Berühren sich Humboldtstraße und Lärchenweg direkt?	f

Map 2, Name Set 2

Table B.8: The questions used for Map 2 with name set 2.

Question	**Question (MAP2, SET2)**	**Correct Answer**
1	Kreuzen sich Dorfstraße und Bergstraße?	t
2	Sind Amselweg und Bergstraße teilweise oder ganz parallel?	t
3	Wird die Bergstraße oben durch den Kartenrand begrenzt?	t
4	Treffen sich Amselweg und Dorfstraße in einer Ecke?	t
5	Wird die Dorfstraße rechts durch den Kartenrand begrenzt?	t
6	Mündet die Blumenstraße in die Bergstraße ein?	t
7	Sind Hochstraße und Dorfstraße teilweise oder ganz parallel?	t
8	Endet die Hochstraße rechts in einer Sackgasse, wird also nicht durch den Kartenrand begrenzt?	t
9	Kreuzen sich Hochstraße und Bergstraße?	t
10	Berühren sich Hochstraße und Amselweg direkt?	t
11	Kreuzen sich Amselweg und Bergstraße?	f
12	Sind Hochstraße und Amselweg teilweise oder ganz parallel?	f
13	Wird die Blumenstraße rechts durch den Kartenrand begrenzt?	f
14	Kreuzen sich Dorfstraße und Hochstraße?	f
15	Wird die Hochstraße rechts durch den Kartenrand begrenzt?	f
16	Mündet die Bergstraße in den Amselweg ein?	f
17	Sind Dorfstraße und Amselweg parallel?	f
18	Endet die Dorfstraße in einer Sackgasse, wird also nicht durch den Kartenrand begrenzt?	f
19	Kreuzen sich Amselweg und Blumenstraße?	f
20	Berühren sich Hochstraße und Dorfstraße direkt?	f

Map 3, Name Set 1

Table B.9: The questions used for Map 3 with name set 1.

Question	Question (MAP3, SET1)	Correct Answer
1	Kreuzen sich Poststraße und Hegelstraße?	t
2	Sind Lärchenweg und Hegelstraße teilweise oder ganz parallel?	t
3	Wird die Goethestraße oben durch den Kartenrand begrenzt?	t
4	Kreuzen sich Lärchenweg und Humboldtstraße?	t
5	Wird die Hegelstraße links durch den Kartenrand begrenzt?	t
6	Mündet der Lärchenweg in die Goethestraße ein?	t
7	Sind Poststraße und Humboldtstraße teilweise oder ganz parallel?	t
8	Endet die Hegelstraße rechts in einer Sackgasse, wird also nicht durch den Kartenrand begrenzt?	t
9	Kreuzen sich Humboldtstraße und Hegelstraße?	t
10	Berühren sich Lärchenweg und Poststraße direkt?	t
11	Kreuzen sich Poststraße und Humboldtstraße?	f
12	Sind Hegelstraße und Poststraße teilweise oder ganz parallel?	f
13	Wird der Lärchenweg rechts durch den Kartenrand begrenzt?	f
14	Kreuzen sich Hegelstraße und Lärchenweg?	f
15	Wird die Hegelstraße rechts durch den Kartenrand begrenzt?	f
16	Mündet die Goethestraße in die Poststraße ein?	f
17	Sind Humboldtstraße und Lärchenweg teilweise oder ganz parallel?	f
18	Endet die Poststraße oben in einer Sackgasse, wird also nicht durch den Kartenrand begrenzt?	f
19	Kreuzen sich Hegelstraße und Goethestraße?	f
20	Berühren sich Hegelstraße und Lärchenweg direkt?	f

Map 3, Name Set 2

Table B.10: The questions used for Map 3 with name set 2.

Question	Question (MAP3, SET2)	Correct Answer
1	Kreuzen sich Amselweg und Blumenstraße?	t
2	Sind Bergstraße und Blumenstraße teilweise oder ganz parallel?	t
3	Wird die Dorfstraße oben durch den Kartenrand begrenzt?	t
4	Kreuzen sich Bergstraße und Hochstraße?	t
5	Wird die Blumenstraße links durch den Kartenrand begrenzt?	t
6	Mündet die Bergstraße in die Dorfstraße ein?	t
7	Sind Amselweg und Hochstraße teilweise oder ganz parallel?	t
8	Endet die Blumenstraße rechts in einer Sackgasse, wird also nicht durch den Kartenrand begrenzt?	t
9	Kreuzen sich Hochstraße und Blumenstraße?	t
10	Berühren sich Bergstraße und Amselweg direkt?	t
11	Kreuzen sich Amselweg und Hochstraße?	f
12	Sind Blumenstraße und Amselweg teilweise oder ganz parallel?	f
13	Wird die Bergstraße rechts durch den Kartenrand begrenzt?	f
14	Kreuzen sich Blumenstraße und Bergstraße?	f
15	Wird die Blumenstraße rechts durch den Kartenrand begrenzt?	f
16	Mündet die Dorfstraße in den Amselweg ein?	f
17	Sind Hochstraße und Bergstraße teilweise oder ganz parallel?	f
18	Endet der Amselweg oben in einer Sackgasse, wird also nicht durch den Kartenrand begrenzt?	f
19	Kreuzen sich Blumenstraße und Dorfstraße?	f
20	Berühren sich Blumenstraße und Bergstraße direkt?	f

B.6 Questionnaire Used in Experiment 2

Systemevaluation

Einleitung: Bitte beantworten Sie die folgenden Fragen, in dem Sie eine Zahl von 1 bis 5 nennen. 1 bedeutet dabei, dass Sie vollkommen einverstanden sind, 5, dass Sie überhaupt nicht einverstanden sind.

- Die haptische Darstellung hilft beim Verstehen der Karte. Antwort: ______

- Taktile Karten sprachlich zu unterstützen ist auf jeden Fall hilfreich. Antwort: ______

- Die gegebenen sprachlichen Äußerungen waren hilfreich. Antwort: ______

  -> Falls nein, welche nicht?

- Ich hätte gerne noch weitere Assistenzen erhalten. Antwort: ______

  -> Falls ja, welche?

- Die Sprache kam immer zum richtigen Zeitpunkt. . Antwort: ______

- Ich fand die Assistenz für Straßen genau so hilfreich wie für den Baum und die Gebäude. Antwort: ______

- Die Aufgabe, die Karte im Puzzle zusammenzusetzen, war verständlich. Antwort ______

Figure B.3: The questionnaire used in Experiment 2.

B.7 Interview Guideline for Experiment 2

1. Erinnern sie sich bitte bitte zurück an den Durchlauf, in dem Sie die Karte mit der erweiterten sprachlichen Assistenz erfühlt hatten, also der Bedingung, in der mehr Informationen zu der Karte gegeben wurden. Wir wüssten gerne, wie sie mit der Aufgabe umgegangen sind. Versuchen sie bitte sich daran zu erinnern, was sie gedacht haben und sie vorgegangen sind, während sie die Karte erfühlt haben. Sprechen Sie einfach alle Gedanken und Strategien, an die Sie sich erinnern, laut aus.
2. (nach erfolgtem Bericht) Welche Rolle haben die sprachlichen Informationen bei ihrer Exploration gespielt? Versuchen sie bitte sich genau zu erinnern, wie sie mit den sprachlichen Informationen umgegangen bist, die sie gehört haben. Uns interessieren alle Gedanken, die mit den sprachlichen Informationen zusammen hängen.
3. (nach erfolgtem Bericht) Es gab sprachliche Informationen zu Straßen. Mich interessiert, welche Gedanken sie zu den Informationen zu Straßen haben.

4. (nach erfolgtem Bericht) Weiter gab es noch sprachliche Informationen zu Gebäuden, mich interessiert, welche Gedanken sie zu diesen Informationen haben.

B.8 Translation of Example Interview Transcript, Experiment 2

The following is a translation of a transcript of an interview conducted with a congenitally blind participant after Experiment 2. Square brackets indicate comments of the author. Dots ('...') without brackets indicate longer pauses.

Experimenter:
I would like to know how you solved the task. Please try to remember what you thought and how you proceeded when you explored the map.

Participant:
All maps?

Experimenter:
Mostly, the last one [this was the one with extended assistance, referred to as 'the second' by the participant in the following; author's comment], but some strategies might stay equally. Just talk about all the thoughts and strategies that you remember.

Participant:
Well, I tried to solve the task systematically, ähm, to begin generally, I felt like ... but maybe that is just me, but I had the feeling that I was much better at the second time [condition, extended assistance]. Well, generally the full task, beginning with the acquisition of the maps, in the first map, there were some things that I did not find, in the second map I found all the buildings and all the streets and all characteristics. That was similar in the first map, but there some buildings were missing, for instance the oak tree, the museum, and the school. I think I found three of five buildings, which were on the map. And, of course, I solved the task systematically. I firstly looked at the map: which buildings or streets are there, concerning the course of streets and their major axes and minor axes, which affect the complete map. For example, if one sees the second map, there are always streets that—ähm, in the first map this was Hochstraße in combination with Bergstraße, that was an axis through pretty much the complete map. And for the second map this was Poststraße and, ... Ähm, back to the system, yes I looked at the course of the streets: Which street crosses which other ... to get an overview. Where are dead ends or, correspondingly, which streets just go from the upper to the lower map frame. That would on the second map be this one street, Hegelstraße, for example. Now then, as you walk outside as a blind man you should always try to find distinctive marks, salient points. I managed that in the second map. There were four buildings; though I did not remember which was on which side, but I knew at which streets they were and which were facing each other, too. When I was done with cognizing the streets I just looked what is to the right and left of this street, what are there for oak trees and where are they. Firstly, at which street are they and from which direction are they on which

side. On the second [condition], for example, the university was close to Poststraße, I found that pretty much first of all. And then there were also areas where I thought, 'OK, you know that that is a street', and 'I know which street that is, but whether there are buildings?' Then I just circled with the pen above that street and then there was nothing and then time was already over and the questions that I answered were posed. Yes, already with the first [condition, author's comment] and with the second my imagination has really helped, as I knew already a little, where is which street and how do they run. I also had to presume many a thing, because it is like—when I had a look—I mean, when I knew a little where which street is and how they run ... I also had to assume some things, because it is not that easy. But that is not a problem, if I had had a look a second or a third time, I would have known it, I suppose ... I am certain! Well, and then came the part with the jigsaw puzzle [the recognition task]. So to speak, you got a map, a tactile one and considered which part would fit to it. I simply tried to remind myself [original German phrase: 'mir ins Gedächtnis zu rufen'] of the course of the streets, that is, some really salient points. For the second, that was ... for the first [condition] I guessed, for the second was the already mentioned Hegelstraße a very large help because the lower right part—well I don't know—the lower right quadrant, can you call it like that?

Experimenter:
Yes!

Participant:
And then I said, so to speak, I compared the map parts with each other. There were several to select from. And there was just one part where the street went form the very top to the very bottom.

Experimenter:
I doubt that, but I think that which you describe is a good strategy.

Participant:
Well, so, the one street went from the top to the bottom! [with certainty, author's comment]

Experimenter:
Yes, but there were more parts than one parts at the jigsaw puzzle where the street went from top to bottom.

Participant:
Ah, well, that might be that there were more parts in which the street went from to to bottom. That might be. But I tried to adopt a strategy in which I tried to remember properties. That might be that there were more parts, I don't want to make a commitment on that now! Well, this are the thoughts I have with respect to that.

Experimenter:
Which role did the verbal information have when exploring the map. Please try to remember how you used the verbal information, that you listened to. We are interested in all thoughts related to that.

Participant:
Well, this was a fancy thing [original German phrase ‘eine feine Sache’]. It really has helped me for orientation, especially for the second [condition]. Ähm, I remember an episode [original German phrase: ‘eine Episode’] where I was in a street and then I just went with the pen to the right and to the left repeatedly and in the indentation [original German phrase: ‘Einbuchtung’], so to speak. Then there came ‘now, the building is to the left and to the right is now that.’ That was really super for orientation and, furthermore, to remember streets.

Experimenter:
There were different kinds of verbal information. On the one hand for streets and on the other for buildings. To begin with, I am interested which thoughts you have with respect to the information for the streets. [. . . , some non-relevant clarification is ommited] I mean, you already gave a lot of information. Is there something to add?

Participant:
Well, it is similar as with the buildings—I found that what has been said good. What I found cool as well, was what has been said for crossings and the further course [original phrase: ‘weiteren Verlauf’] of the streets. That not only was said ‘that is now that and that crossing’, but more like: ‘if you go now further here, then this street runs into another street, or it is a dead end, or it runs from the upper to the lower map frame’. That is important. For the future, if one does it like it is now [implement the system], because if a sighted uses a city map or an atlas, these are maps as well [sic!]. And now when a route from A to B extends a page, then I need a new map and I have to open the new page. Therefore, this was actually good.

Experimenter:
There were, additionally information for buildings, which thoughts to you have with respect to that? Do you wish to add something?

Participant:
No, actually, for the second [condition], there was something. . . . I don’t remember exactly. I think it was somehow like a building was diagonal to a street? I think that was something with the university and the oak, maybe?

Experimenter:
There was information that buildings are left and right below or above intersections.

Participant:
Exactly, that is it. Yes, when I consider to really walk there and I get such information, that is not bad. Well, I don’t have really more to say.

Experimenter:
Thank you!

B.9 Questionnaire Used in Experiment 3

1. Die haptische Karte ist verständlich.
trifft vollkommen zu □ □ □ □ □ trifft überhaupt nicht zu

2. Solche Karten sprachlich zu unterstützen ist auf jeden Fall hilfreich.
trifft vollkommen zu □ □ □ □ □ trifft überhaupt nicht zu

3. Die gegebenen sprachlichen Äußerungen waren verständlich.
trifft vollkommen zu □ □ □ □ □ trifft überhaupt nicht zu

4. Ich wusste immer genau, was mit den sprachlichen Informationen gemeint ist.
trifft vollkommen zu □ □ □ □ □ trifft überhaupt nicht zu

5. Die gegebenen sprachlichen Äußerungen waren hilfreich.
trifft vollkommen zu □ □ □ □ □ trifft überhaupt nicht zu

Falls nicht, welche?

6. Ich hätte gerne noch weitere sprachliche Informationen erhalten.
trifft vollkommen zu □ □ □ □ □ trifft überhaupt nicht zu

Falls ja, welche?

7. Die Sprache kam immer zum richtigen Zeitpunkt.
trifft vollkommen zu □ □ □ □ □ trifft überhaupt nicht zu

8. Ich kann mich in einer unbekannten Stadt gut mit einem Stadtplan orientieren.
trifft vollkommen zu □ □ □ □ □ trifft überhaupt nicht zu

9. Auch mit einem Stadtplan verliere ich manchmal die Orientierung, wenn ich eine Stadt nicht so gut kenne.
trifft vollkommen zu □ □ □ □ □ trifft überhaupt nicht zu

Figure B.4: The questionnaire used in Experiment 3.

B.10 Interview Guideline for Experiment 3

1. (während der Beantwortung der Likert-type-Frage 'Die gegebenen sprachlichen Äußerungen waren hilfreich.') Falls nicht, welche?
2. (während der Beantwortung der Likert-type-Frage 'Ich hätte gerne noch weitere sprachliche Informationen erhalten.') Falls ja, welche?
3. Erinnere Dich bitte zurück daran, wie es war, die Karte zu lernen. Wie bist Du vorgegangen?
4. (nach erfolgtem Bericht) War es verständlich, die Karte mit dem System zu lernen?
5. (nach erfolgtem Bericht) Gab es spezielle Probleme mit dem System?
6. (nach erfolgtem Bericht) Hast Du Vorschläge, wie man das System verbessern kann?

Appendix C

Example Referential Net

frame	pfr	ιx map_frame(x)
geometry_is(...)		
parts([pfr1, pfr2, pfr3, pfr4])		
frame_segment	pfr1	ηx has_upper_limit(pt1, x)
geometry_is(...)		ηx has_upper_limit(pt4, x)
part_of(pfr)		ιx top_frame(x)
frame_segment	pfr2	ηx has_right_limit(pt2, x)
geometry_is(...)		ιx right_frame(x)
part_of(pfr)		
frame_segment	pfr3	ηx has_lower_limit(pt3, x)
geometry_is(...)		ηx has_lower_limit(pt4, x)
part_of(pfr)		ιx bottom_frame(x)
frame_segment	pfr4	ηx has_left_limit(pt1, x)
geometry_is(...)		ιx left_frame(x)
part_of(pfr)		
landmark	plm1	Buche
geometry_is(...)		ηx is_left(pt1, x)
		ηx is_right(x, pt1)
		ηx tree(x)

landmark
geometry_is(...)
— plm2 —
Lidl
ηx building(x)
ηx is_above(x, pt5)
ηx is_below(pt5, x)
ηx is_left_above(x, ptco2)
ηx is_left(x, pt4)
ηx is_right_below(ptco2, x)
ηx is_right(pt4, x)

landmark
geometry_is(...)
— plm3 —
Anne-Frank-Schule
ηx building(x)
ηx is_above(x, pt5)
ηx is_below(pt5, x)
ηx is_left_below(ptco2, x)
ηx is_left(pt4, x)
ηx is_right_above(x, ptco2)
ηx is_right(x, pt4)

landmark
geometry_is(...)
— plm4 —
Museum
ηx building(x)
ηx is_above(pt1, x)
ηx is_below(x, pt1)

landmark
geometry_is(...)
— plm5 —
Gedächtniskirche
ηx building(x)
ηx is_above(x, pt2)
ηx is_below(pt2, x)
ηx is_left_above(x, ptco5)
ηx is_left(x, pt4)
ηx is_right_below(ptco5, x)
ηx is_right(pt4, x)

landmark
geometry_is(...)
— plm6 —
Rathaus
ηx building(x)
ηx is_above(x, pt2)
ηx is_below(pt2, x)
ηx is_left_below(ptco5, x)
ηx is_left(pt4, x)
ηx is_right_above(x, ptco5)
ηx is_right(x, pt4)

track
geometry_is(...)
parts([pts11, pts12, pts13, pts14, pts15])
— pt1 —
Bergstraße
ηx has_left_limit(x, pfr4)
ηx has_upper_limit(x, pfr1)
ηx is_in_track_config(x, ptco3)
ηx street(x)

track
geometry_is(...)
parts([pts21, pts22, pts23, pts24, pts25, pts26, pts27])

pt2

Hochstraße
ηx has_right_limit(x, pfr2)
ηx has_upper_left_limit(x, ptco3)
ηx is_in_track_config(x, ptco3)
ηx is_in_track_config(x, ptco4)
ηx is_in_track_config(x, ptco5)
ηx street(x)

track
geometry_is(...)
parts([pts31, pts32, pts33])

pt3

Amselweg
ηx has_lower_limit(x, pfr3)
ηx has_upper_limit(x, ptco1)
ηx is_in_track_config(x, ptco1)
ηx is_in_track_config(x, ptco4)
ηx is_parallel_to(pt4, x)
ηx is_parallel_to(x, pt4)
ηx street(x)

track
geometry_is(...)
parts([pts41, pts42, pts43, pts44, pts45, pts46])

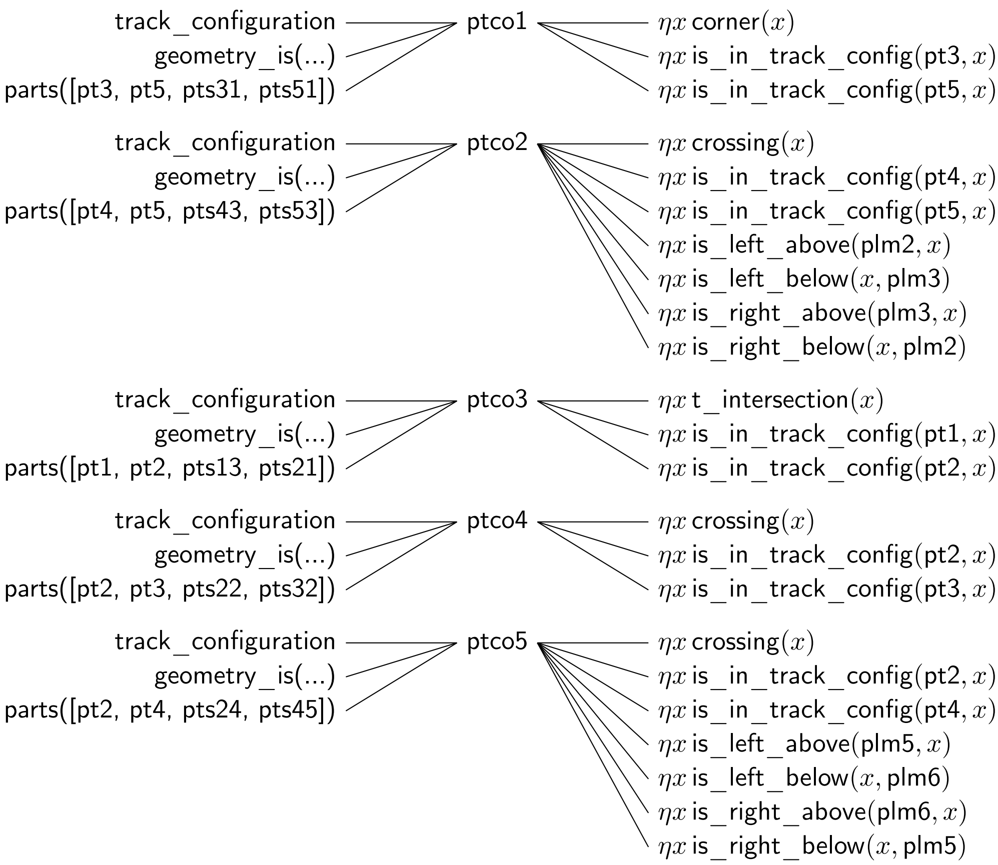
pt4

Blumenstraße
ηx has_lower_limit(x, pfr3)
ηx has_upper_limit(x, pfr1)
ηx is_in_track_config(x, ptco2)
ηx is_in_track_config(x, ptco5)
ηx is_left(plm2, x)
ηx is_left(plm5, x)
ηx is_left(x, plm3)
ηx is_left(x, plm6)
ηx is_parallel_to(pt3, x)
ηx is_parallel_to(x, pt3)
ηx is_right(plm3, x)
ηx is_right(plm6, x)
ηx is_right(x, plm2)
ηx is_right(x, plm5)
ηx street(x)

track
geometry_is(...)
parts([pts51, pts52, pts53, pts54, pts55])

pt5

Dorfstraße
ηx has_left_limit(x, ptco1)
ηx has_right_limit(x, pts55)
ηx is_above(plm2, x)
ηx is_above(plm3, x)
ηx is_below(x, plm2)
ηx is_below(x, plm3)
ηx is_in_track_config(x, ptco1)
ηx is_in_track_config(x, ptco2)
ηx street(x)


track_configuration
geometry_is(...)
parts([pt3, pt5, pts31, pts51])
ptco1
ηx corner(x)
ηx is_in_track_config(pt3, x)
ηx is_in_track_config(pt5, x)
track_configuration
geometry_is(...)
parts([pt4, pt5, pts43, pts53])
ptco2
ηx crossing(x)
ηx is_in_track_config(pt4, x)
ηx is_in_track_config(pt5, x)
ηx is_left_above(plm2, x)
ηx is_left_below(x, plm3)
ηx is_right_above(plm3, x)
ηx is_right_below(x, plm2)
track_configuration
geometry_is(...)
parts([pt1, pt2, pts13, pts21])
ptco3
ηx t_intersection(x)
ηx is_in_track_config(pt1, x)
ηx is_in_track_config(pt2, x)
track_configuration
geometry_is(...)
parts([pt2, pt3, pts22, pts32])
ptco4
ηx crossing(x)
ηx is_in_track_config(pt2, x)
ηx is_in_track_config(pt3, x)
track_configuration
geometry_is(...)
parts([pt2, pt4, pts24, pts45])
ptco5
ηx crossing(x)
ηx is_in_track_config(pt2, x)
ηx is_in_track_config(pt4, x)
ηx is_left_above(plm5, x)
ηx is_left_below(x, plm6)
ηx is_right_above(plm6, x)
ηx is_right_below(x, plm5)

track_segment geometry_is(...) part_of(pt1)	pts11	ηx has_upper_limit(x, pfr1) ηx is_part_of(x, pt1)
track_segment geometry_is(...) part_of(pt1)	pts12	ηx is_left(x, plm1) ηx is_part_of(x, pt1) ηx is_right(plm1, x)
track_segment geometry_is(...) part_of(ptco3) part_of(pt1)	pts13	ηx is_part_of(x, pt1)
track_segment geometry_is(...) part_of(pt1)	pts14	ηx is_above(x, plm4) ηx is_below(plm4, x) ηx is_part_of(x, pt1)
track_segment geometry_is(...) part_of(pt1)	pts15	ηx has_left_limit(x, pfr4) ηx is_part_of(x, pt1)
track_segment geometry_is(...) part_of(ptco3) part_of(pt2)	pts21	ηx has_upper_left_limit(x, ptco3) ηx is_part_of(x, pt2)
track_segment geometry_is(...) part_of(ptco4) part_of(pt2)	pts22	ηx is_part_of(x, pt2)
track_segment geometry_is(...) part_of(pt2)	pts23	ηx is_above(plm5, x) ηx is_below(x, plm5) ηx is_part_of(x, pt2)

track_segment geometry_is(...) part_of(ptco5) part_of(pt2)	pts24	ηx is_part_of(x, pt2)
track_segment geometry_is(...) part_of(pt2)	pts25	ηx is_above(plm6, x) ηx is_below(x, plm6) ηx is_part_of(x, pt2)
track_segment geometry_is(...) part_of(pt2)	pts26	ηx has_right_limit(x, pfr2) ηx is_part_of(x, pt2)
track_segment geometry_is(...) part_of(pt2)	pts27	ηx is_above(plm5, x) ηx is_above(plm6, x) ηx is_below(x, plm5) ηx is_below(x, plm6) ηx is_parallel_to(x, pt5) ηx is_part_of(x, pt2)
track_segment geometry_is(...) part_of(ptco1) part_of(pt3)	pts31	ηx has_upper_limit(x, ptco1) ηx is_part_of(x, pt3)
track_segment geometry_is(...) part_of(ptco4) part_of(pt3)	pts32	ηx is_part_of(x, pt3)
track_segment geometry_is(...) part_of(pt3)	pts33	ηx has_lower_limit(x, pfr3) ηx is_part_of(x, pt3)
track_segment geometry_is(...) part_of(pt4)	pts41	ηx has_upper_limit(x, pfr1) ηx is_part_of(x, pt4)
track_segment geometry_is(...) part_of(pt4)	pts42	ηx is_between(x, plm2, plm3) ηx is_part_of(x, pt4)
track_segment geometry_is(...) part_of(ptco2) part_of(pt4)	pts43	ηx is_part_of(x, pt4)

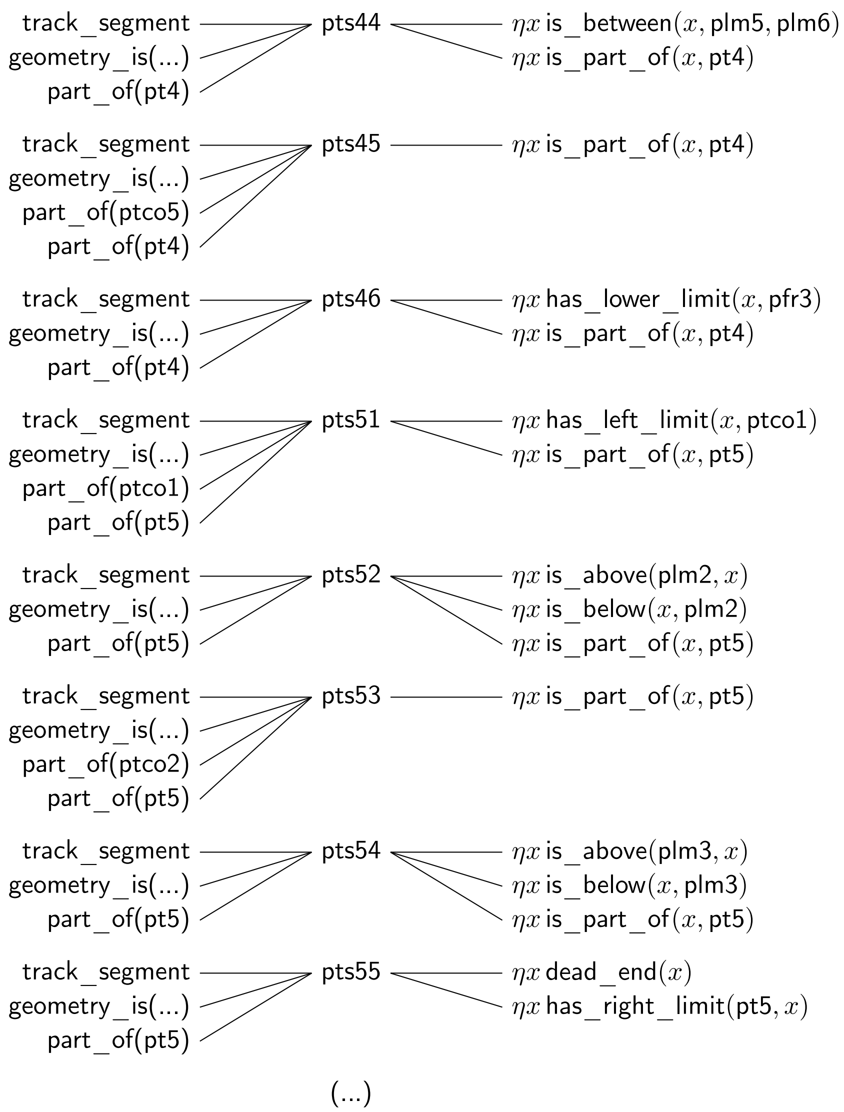


Figure C.1: A referential net representing Map 1 with name set 2.

Appendix D

Example Map

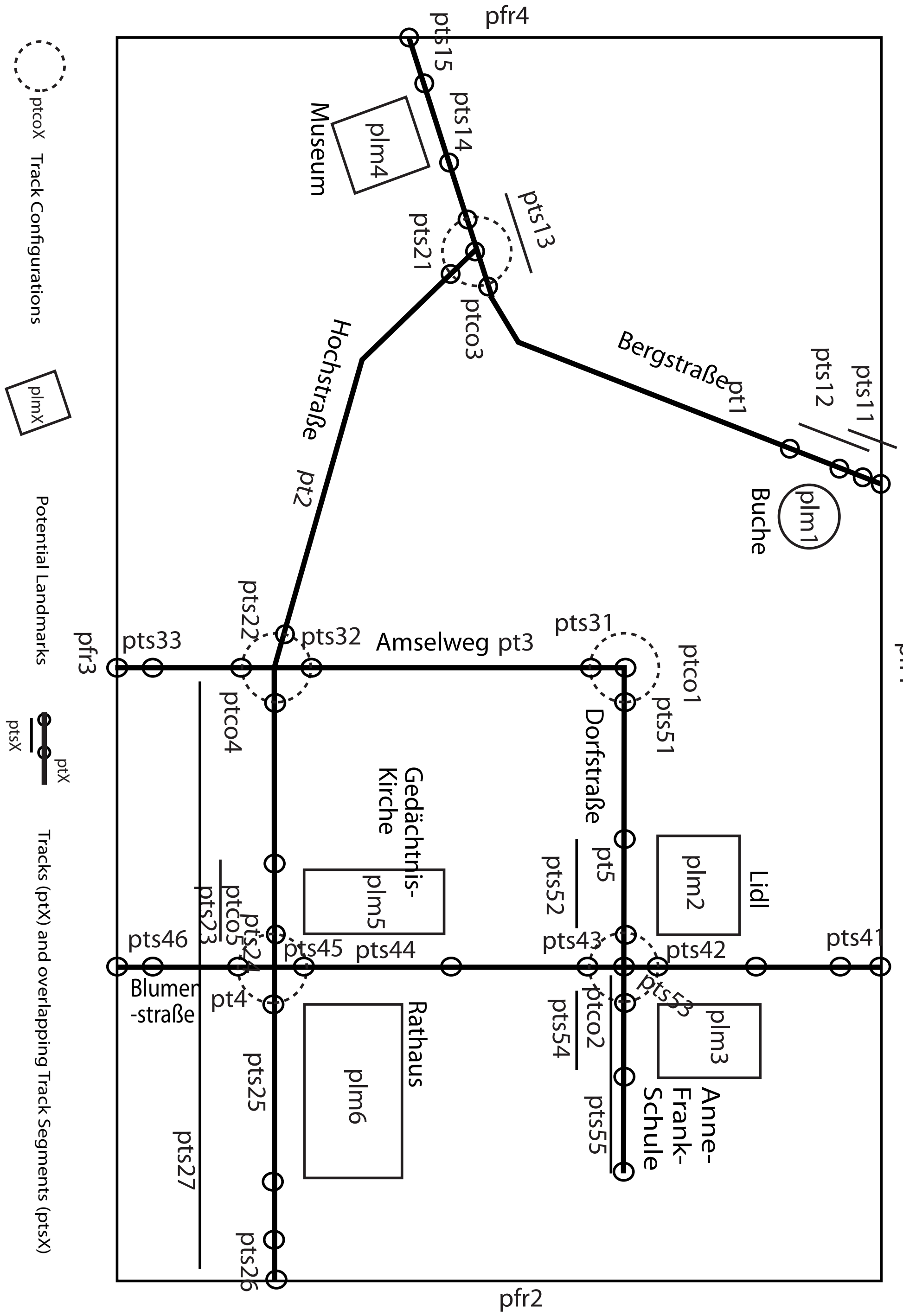


Figure D.1: Map 1 used in the experiments and in the prototype implementation with map entities visualized.

Bibliography

Acartürk, C. (2009). *Multimodal comprehension of graph-text constellations: An information processing perspective.* Unpublished doctoral dissertation, Universität Hamburg, Hamburg.

André, E., Herzog, G., & Rist, T. (1988). On the simultaneous interpretation of real world image sequences and their natural language description: The system soccer. In *Proceedings of the 8th European conference on articifial intelligence* (pp. 449–454).

Andrews, S. (1983). Spatial cognition through tactual maps. In (pp. 30–40).

Appleyard, D. (1969). Why buildings are known: A predictive tool for architects and planners. *Environment and Behavior*, *1*(2), 131–156.

Aragones, J., & Arredondo, J. (1985). Structure of urban cognitive maps. *Journal of Environmental Psychology*, *5*(2), 197–212.

Atkin, A. (2010). Peirce's theory of signs. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy.* Retrieved from `http://plato.stanford.edu/archives/win2010/entries/peirce-semiotics/`

Avraamides, M., Loomis, J., Klatzky, R., & Golledge, R. (2004). Functional equivalence of spatial representations derived from vision and language: evidence from allocentric judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 801.

Baumann, T., Atterer, M., & Schlangen, D. (2009). Assessing and improving the performance of speech recognition for incremental systems. In *Proceedings of human language technologies: The 2009 annual conference of the north american chapter of the association for computational linguistics* (pp. 380–388). Boulder, CO, USA.

Baumann, T., & Schlangen, D. (2012). INPRO_iSS: a component for just-in-time incremental speech synthesis. In *Proceedings of ACL 2012.* Jeju Island, South Korea.

Beck, R. J., & Wood, D. (1976). Cognitive transformation of information from urban geographic fields to mental maps. *Environment and Behavior*, *8*(2), 199–238.

Becker, T., Kilger, A., Lopez, P., & Poller, P. (2000). The verbmobil generation component VM-GECO. In E. Wahlster (Ed.), *Verbmobil: Foundations of speech-to-speech translation* (pp. 481–496). Berlin, Heidelberg, New York: Springer.

Bell, S., & Archibald, J. (2011). Sketch mapping and geographic knowledge: What role for drawing ability? In J. Wang, K. Broelemann, M. Chipofya, A. Schwering, & J. Wallgrün (Eds.), *Understanding and processing sketch maps – proceedings of the COSIT 2011 workshop* (Vol. 42). Heidelberg: AKA.

Benotti, L., & Bertoa, N. (2011). Content determination through planning for flexible game tutorials. *Advances in Artificial Intelligence*, 345–356.

Bentzen, B. (1972). Production and testing of an orientation and travel map for visually handicapped persons. *The New Outlook for the Blind*, *66*.

Bérla, E. P. (1981). Tactile scanning and memory for a spatial display by blind students. *The Journal of Special Education*, *15*(3), 341–350.

Bérla, E. P., Butterfield, L., & Murr, M. (1976). Tactual reading of political maps by blind students: A videomatic behavioral analysis. *The Journal of Special Education*, *10*(3), 265–276.

Bernsen, N. (1994). Foundations of multimodal representations: a taxonomy of representational modalities. *Interacting with computers*, *6*(4), 347–371.

Bigelow, A. E. (1996). Blind and sighted children's spatial knowledge of their home environments. *International Journal of Behavioral Development*, *19*(4), 797 –816.

Billinghurst, M., & Weghorst, S. (1995). The use of sketch maps to measure cognitive maps of virtual environments. In *Proceedings of the virtual reality annual international symposium* (pp. 40 – 47).

Blades, M. (1990). The reliability of data collected from sketch maps. *Journal of Environmental Psychology*, *10*(4), 327–339.

Blades, M., Ungar, S., & Spencer, C. (1999). Map use by adults with visual impairments. *The Professional Geographer*, *51*(4), 539–553.

Bradley, J. (1958). Complete counterbalancing of immediate sequential effects in a latin square design. *Journal of the American Statistical Association*, 525–528.

Brambring, M., & Weber, C. (1981). Taktile, verbale und motorische Informationen zur geographischen Orientierung Blinder. *Zeitschrift für experimentelle und angewandte Psychologie*, *28*(1), 23–37.

Brunyé, T. T., & Taylor, H. A. (2008a). Extended experience benefits spatial mental model development with route but not survey descriptions. *Acta Psychologica*, *127*(2), 340–354.

Brunyé, T. T., & Taylor, H. A. (2008b). Working memory in developing and applying mental models from spatial descriptions. *Journal of Memory and Language*, *58*(3), 701–729.

Buzzi, M. C., Buzzi, M., Leporini, B., & Martusciello, L. (2011). Making visual maps accessible to the blind. *Universal Access in Human-Computer Interaction. Users Diversity*, 271–280.

Carlson-Radvansky, L. A., & Irwin, D. E. (1993). Frames of reference in vision and language: Where is above? *Cognition*, *46*(3), 223–244.

Carstensen, K. U., Ebert, C., Ebert, C., Jekat, S., Langer, H., & Klabunde, R. (Eds.). (2009). *Computerlinguistik und Sprachtechnologie: Eine Einführung*. Springer.

Cattaneo, Z., & Vecchi, T. (2011). *Blind vision: The neuroscience of visual impairment*. Cambridge, MA: MIT Press.

Cattaneo, Z., Vecchi, T., Cornoldi, C., Mammarella, I., Bonino, D., Ricciardi, E., & Pietrini, P. (2008). Imagery and spatial processes in blindness and visual impairment. *Neuroscience & Biobehavioral Reviews*, *32*(8), 1346–1360.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Colwell, C., Petrie, H., Kornbrot, D., Hardwick, A., & Furner, S. (1998a). Haptic virtual reality for blind computer users. In *Proceedings of the third international ACM conference on assistive technologies* (pp. 92–99). ACM.

Colwell, C., Petrie, H., Kornbrot, D., Hardwick, A., & Furner, S. (1998b). Use of a haptic device by blind and sighted people: perception of virtual textures and objects. In *Proceedings of the third congress for technology for inclusive design and equality.* Helsinki, Finland.

Curiel, J. M., & Radvansky, G. A. (1998). Mental organization of maps. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 202–214.

Curiel, J. M., & Radvansky, G. A. (2004). The accuracy of spatial information from temporally and spatially organized mental maps. *Psychonomic Bulletin and Review*, *11*, 314–319.

Dahlback, N., Jonsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies–why and how. *Knowledge-based Systems*, *6*(4), 258–266.

Daniel, M., Ariane, T., Manghi, E., & Denis, M. (2003). Testing the value of route directions through navigational performance. *Spatial Cognition & Computation*, *3*(4), 269 – 289.

Davenport, E. C., & El-Sanhurry, N. A. (1991). Phi/Phimax: review and synthesis. *Educational and Psychological Measurement*, *51*(4), 821–828.

de Almeida (Vasconcellos), R. A., & Tsuji, B. (2005). Interactive mapping for people who are blind or visually impaired. In D. R. Fraser Taylor (Ed.), *Modern cartography series* (Vol. 4, pp. 411–431). Academic Press.

De Felice, F., Renna, F., Attolico, G., & Distante, A. (2007). A haptic/acoustic application to allow blind the access to spatial information. In *Proceedings of world haptics 2007* (pp. 310 – 315). Tsukuba, Japan.

Denis, M. (1997). The description of routes: A cognitive approach to the production of spatial discourse. *Cahiers de psychologie cognitive*, *16*(4), 409–458.

Denis, M., & Cocude, M. (1989). Scanning visual images generated from verbal descriptions. *European Journal of Cognitive Psychology*, *1*(4), 293–307.

Denis, M., & Zimmer, H. (1992). Analog properties of cognitive maps constructed from verbal descriptions. *Psychological Research*, *54*(4), 286–298.

De Smedt, K., Horacek, H., & Zock, M. (1996). Architectures for natural language generation: Problems and perspectives. *Trends in Natural Language Generation An Artificial Intelligence Perspective*, 17–46.

Dethlefs, N. (2011). The Bremen system for the GIVE-2.5 challenge. In *Proceedings of the 13th European workshop on natural language generation* (pp. 284–289). Nancy, France.

Downs, R., & Stea, D. (1977). *Maps in minds: Reflections on cognitive mapping.* New York: Harper & Row.

Eimer, M. (2004). Multisensory integration: How visual experience shapes spatial perception. *Current Biology*, *14*(3), 115–117.

Eschenbach, C. (1988). *SRL als Rahmen eines textverarbeitenden Systems* (Tech. Rep.). Universität Hamburg. Retrieved from `http://www.informatik.uni-hamburg.de/WSV/pub/Eschenbach-1988-SRL.pdf`

Eschenbach, C. (1999). Geometric structures of frames of reference and natural language semantics. *Spatial Cognition and Computation*, *1*(4), 329–348.

Espinosa, M., Ungar, S., Ochaita, E., Blades, M., & Spencer, C. (1998). Comparing methods for introducing blind and visually impaired people to unfamiliar urban environments. *Journal of Environmental Psychology*, *18*, 277–287.

Evans, D. G., & Blenkhorn, P. (1995). Access to graphical information for blind people through speech and touch. In I. P. Porrero & R. P. de la Bellacasa Alberola (Eds.), *Proceedings of the 2nd tide congress* (pp. 298–305). Paris, France.

Evans, G. W., & Pezdek, K. (1980). Cognitive mapping: Knowledge of real-world distance and location information. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(1), 13.

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London, Thousand Oaks, CA, New Delhi: Sage.

Field, A., & Hole, G. (2003). *How to design and report experiments* (2nd ed.). Los Angeles, London, Neu Delhi, Singapore, Washington D.C.: Sage.

Fletcher, J. (1980). Spatial representation in blind children. 1: Development compared to sighted children. *Journal of Visual Impairment and Blindness*, *74*(12), 381–85.

Foulke, E. (1982). Perception, cognition, and the mobility of blind pedestrians. In M. Potegal (Ed.), *Spatial abilities: Development and physiological foundations* (pp. 55–76). New York: Academic Press.

Fricke, E. (2008). *Grundlagen einer multimodalen Grammatik des Deutschen.* Habilitationsschrift, Europa-Universität Viadrina, Frankfurt (Oder), Germany.

Garoufi, K., & Koller, A. (2010). Automated planning for situated natural language generation. In *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 1573–1582). Uppsala, Sweden.

Garoufi, K., & Koller, A. (2011). Combining symbolic and corpus-based approaches for the generation of successful referring expressions. In *Proceedings of the 13th European workshop on natural language generation* (pp. 121–131). Nancy, France.

Giudice, N. A. (2004). *Navigating novel environments: A comparison of verbal and visual learning.* Unpublished doctoral dissertation, University of Minnesota.

Giudice, N. A. (2006). Wayfinding without vision: Learning real and virtual environments using dynamically-updated verbal descriptions. In *Proceedings of the conference on assistive technologies for vision and hearing impairment.* Kufstein, Austria.

Giudice, N. A., Bakdash, J. Z., & Legge, G. E. (2007). Wayfinding with words: spatial learning and navigation using dynamically updated verbal descriptions. *Psychological Research*, *71*(3), 347–358.

Giudice, N. A., Bakdash, J. Z., Legge, G. E., & Roy, R. (2010). Spatial learning and navigation using a virtual verbal display. *ACM Transactions on Applied Perception*, *7*(1), 1–22.

Giudice, N. A., Betty, M. R., & Loomis, J. M. (2011). Functional equivalence of spatial images from touch and vision: Evidence from spatial updating in blind and sighted individuals. *Learning, Memory*, *37*(3), 621–634.

Giudice, N. A., & Legge, G. E. (2008). Blind navigation and the role of technology.

In A. Helal, M. Mokthari, & B. Abdulrazak (Eds.), *The engineering handbook of smart technology for aging, disability, and independence* (pp. 479–500). Hoboken, NJ, USA: Jon Wiley.

Golledge, R. G. (1976). Methods and methodological issues in environmental cognition research. In G. T. Moore (Ed.), *Environmental knowing: theories, research, and methods* (pp. 300–313). Sroudsburg, Dowden: Hutchinson & Ross.

Golledge, R. G. (1993). Geography and the disabled: a survey with special reference to vision impaired and blind populations. *Transactions of the Institute of British Geographers*, *18*(1), 63–85.

Golledge, R. G. (1999). Human wayfinding and cognitive maps. In R. G. Golledge (Ed.), *Wayfinding behavior: Cognitive mapping and other spatial processes* (pp. 5–45). Baltimore, MD, USA: Johns Hopkins University Press.

Golledge, R. G., Dougherty, V., & Bell, S. (1995). Acquiring spatial knowledge: Survey versus route-based knowledge in unfamiliar environments. *Annals of the Association of American Geographers*, *85*(1), 134–158.

Golledge, R. G., Klatzky, R. L., & Loomis, J. M. (1996). Cognitive mapping and wayfinding by adults without vision. In J. Portugali (Ed.), *The construction of cognitive maps* (pp. 215–246). Doordrecht: Kluwer.

Golledge, R. G., Rice, M., & Jacobson, R. D. (2005). A commentary on the use of touch for accessing on-screen spatial representations: The process of experiencing haptic maps and graphics. *The Professional Geographer*, *57*(3), 339–349.

Goodman, N. (1976). *Languages of art: An approach to a theory of symbols.* Indianapolis, IN: Hackett.

Guhe, M. (2007). *Incremental conceptualization for language production.* Mahwah, NJ: Lawrence Erlbaum Associates.

Guhe, M., Habel, C., & Tappe, H. (2000). Incremental event conceptualization and natural language generation in monitoring environments. In *Proceedings of the first international conference on natural language generation* (Vol. 14, pp. 85 – 92). Mitzpe Ramon, Israel.

Guhe, M., Habel, C., & Tschander, L. (2003). Describing motion events: Incremental representations for incremental processing. *Proceedings of the 5th International Workshop on Computational Semantics (IWCS-5)*, 410–424.

Guhe, M., Habel, C., & Tschander, L. (2004). Incremental generation of interconnected preverbal messages. In T. Pechmann & C. Habel (Eds.), *Multidisciplinary approaches to language production* (pp. 7–52). Berlin, New York: De Gruyter.

Habel, C. (1982). Referential nets with attributes. In *Proceedings of the 9th conference on computational linguistics* (pp. 101–106). Praha, Czechoslovakia.

Habel, C. (1986). *Prinzipien der Referentialität.* Berlin, Heidelberg, New York: Springer.

Habel, C. (1987). Cognitive linguistics: The processing of spatial concepts. *T. A. Informations (Bulletin semestriel de l'ATALA, Association pour le traitement automatique du langage)*, *28*(2), 21–56.

Habel, C. (1996). Representations as basis of cognitive processes. In G. Görz & S. Hölldobler (Eds.), *Ki-96: Advances in artificial intelligence: 20th annual german con-*

ference on artificial intelligence, Lecture Notes in Artificial Intelligence (Vol. 1137, pp. 99–101). Berlin: Springer.

Habel, C. (2003a). Incremental generation of multimodal route instructions. In *Papers from the 2003 AAAI spring symposium on natural language generation in spoken and written dialogue* (pp. 44–51). Stanford, CA, USA.

Habel, C. (2003b). Representational commitment in maps. In M. Duckham, M. F. Goodchild, & M. F. Worboys (Eds.), *Foundations of geographic information science* (pp. 69 – 93). London: Taylor & Francis.

Habel, C., Kerzel, M., & Lohmann, K. (2010). Verbal assistance in tactile-map explorations: A case for visual representations and reasoning. In *Proceedings of the AAAI workshop on visual representations and reasoning 2010.* Menlo Park, CA.

Habel, C., Pribbenow, S., & Simmons, G. (1995). Partonomies and depictions: A hybrid approach. In J. Glasgow, H. Narayanan, & B. Chandrasekaran (Eds.), *Diagrammatic reasoning: Cognitive and computational perspectives* (pp. 627–653). Cambridge, MA: MIT Press.

Habel, C., & Tappe, H. (1999). Processes of segmentation and linearization in describing events. In R. Klabude & C. von Stutterheim (Eds.), *Processes in language production* (pp. 117–153). Opladen: Westdeutscher Verlag.

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence*, *34*(2), 151–176.

Herman, J. F., Herman, T. G., & Chatman, S. P. (1983). Constructing cognitive maps from partial information: A demonstration study with congenitally blind subjects. *Journal of Visual Impairment and Blindness*, *77*(5), 195–198.

Herzog, G., & Retz-Schmidt, G. (1990). Das System SOCCER: Simultane Interpretation und natürlichsprachliche Beschreibung zeitveranderlicher Szenen. *Sport und Informatik*, 95–119.

Herzog, G., & Wazinski, P. (1994). Visual translator: Linking perceptions and natural language descriptions. *Artificial Intelligence Review*, *8*(2), 175–187.

Hovy, E. (1993). Automated discourse generation using discourse structure relations. *Artificial Intelligence*, *63*(1–2), 341–385.

Howard, R. B., Chase, S. D., & Rothman, M. (1973). An analysis of four measures of cognitive maps. In W. F. Preiser (Ed.), *Environmental design research* (Vol. 1, pp. 254–264). Stroudsburg, PA: Dowden.

Hudson, M. (1998). The michigan state university talking tactile map project: Advancement throught collaboration. In *Proceedings of center on disabilities technology and persons with disabilities conference 1998 (CSUN 1998).*

Hund, A. M., Haney, K. H., & Seanor, B. D. (2008). The role of recipient perspective in giving and following wayfinding directions. *Applied Cognitive Psychology*, *22*(7), 896–916.

Huttenlocher, J., & Strauss, S. (1968). Comprehension and a statement's relation to the situation it describes. *Journal of Verbal Learning and Verbal Behavior*, *7*(2), 300–304.

Ishikawa, T., & Montello, D. (2006). Spatial knowledge acquisition from direct experience in the environment: Individual differences in the development of metric knowledge and the integration of separately learned places. *Cognitive Psychology*, *52*(2), 93–129.

Jacobson, R. D. (1992). Spatial cognition through tactile mapping. *Swansea Geographer*, *29*, 79–88.

Jacobson, R. D. (1996). Talking tactile maps and environmental audio beacons: An orientation and mobility development tool for visually impaired people. In *Proceedings of maps and diagrams for blind and visually impaired people: Needs, solutions, and developments* (pp. 21–25). Ljubjiana, Slovenia.

Jacobson, R. D. (1998a). Cognitive mapping without sight: Four preliminary studies of spatial learning. *Journal of Environmental Psychology*, *18*, 289–306.

Jacobson, R. D. (1998b). Navigating maps with little or no sight: An audio-tactile approach. In *Proceedings of the workshop on content visualization and intermedia representations (CVIR)* (pp. 95 – 102). Montreal.

Jacobson, R. D., & Kitchin, R. M. (1997). GIS and people with visual impairments or blindness: Exploring the potential for education, orientation, and navigation. *Transactions in GIS*, *2*, 315–332.

Kaklanis, N., Votis, K., Moschonas, P., & Tzovaras, D. (2011). HapticRiaMaps: towards interactive exploration of web world maps for the visually impaired. In *Proceedings of the international cross-disciplinary conference on web accessibility.* Hyderabad, India.

Kerzel, M., & Habel, C. (2011). Monitoring and describing events for virtual-environment tactile-map exploration. In A. Galton, M. F. Worboys, & M. Duckham (Eds.), *Proceedings of workshop on identifying objects, processes and events* (pp. 13–18). Belfast, ME, USA.

Kerzel, M., & Habel, C. (2012). Ereigniserkennung während der Exploration audio-taktiler Karten. In *Proceedings of Mensch & Computer 2012.* Konstanz, Germany: Oldenbourg Verlag.

Kipp, M. (2001). A generic annotation tool for multimodal dialogue. In *Proceedings of the 7th European conference on speech communication and technology (Eurospeech 2001)* (pp. 1367–1370). Aalborg, Danmark.

Kitchin, R. M. (1996). Methodological convergence in cognitive mapping research: Investigating configurational knowledge. *Journal of Environmental Psychology*, *16*(3), 163–185.

Kitchin, R. M. (2000). Collecting and analysing cognitive mapping data. In R. Kitchin & S. Freudschuh (Eds.), *Cognitive mapping: past, present, and future.* (pp. 9–24). London: Routledge.

Kitchin, R. M., & Blades, M. (2002). *The cognition of geographic space.* London, New York: I. B. Tauris.

Kitchin, R. M., & Jacobson, R. D. (1997). Techniques to collect and analyze the cognitive map knowledge of persons with visual impairment or blindness: Issues of validity. *Journal of Visual Impairment and Blindness*, *91*(4), 360–376.

Klatzky, R. L., Golledge, R. G., Loomis, J. M., & Cicinelli, J. G. (1995). Performance of blind and sighted persons on spatial tasks. *Journal of Visual Impairment and Blindness*, *89*(1).

Klatzky, R. L., Loomis, J. M., Golledge, R. G., & Cicinelli, J. G. (1990). Acquisition of route and survey knowledge in the absence of vision. *Journal of Motor Behavior*, *22*(1), 19–43.

Kosslyn, S. M. (1973). Scanning visual images: Some structural implications. *Attention, Perception, & Psychophysics*, *14*(1), 90–94.

Kostopoulos, K., Moustakas, K., Tzovaras, D., & Nikolakis, G. (2007). Haptic access to conventional 2D maps for the visually impaired. *Journal on Multimodal User Interfaces*, *1*(2), 13–19.

Krueger, M. W., & Gilden, D. (1997). KnowWhere$^{\mathrm{TM}}$: an audio/spatial interface for blind people. In *Proceedings of the fourth international conference on auditory display (ICAD 97)* (pp. 1–4). Palo Alto, CA, USA.

Kuipers, B. (1982). The "map in the head" metaphor. *Environment and Behavior*, *14*(2), 202.

Lahav, O., & Mioduser, D. (2000). Multisensory virtual environment for supporting blind persons' acquisition of spatial cognitive mapping, orientation, and mobility skills. In *Proceedings of the third international conference on disability, virtual reality and associated technologies, ICDVRAT 2000* (pp. 23–25). Alghero, Italy.

Lahav, O., & Mioduser, D. (2003). A blind person's cognitive mapping of new spaces using a haptic virtual environment. *Journal of Research in Special Educational Needs*, *3*(3), 172–177.

Lahav, O., & Mioduser, D. (2004). Blind persons' acquisition of spatial cognitive mapping and orientation skills supported by virtual environment. In *Proceedings of the 5th international conference on disability, virtual reality and associated technologies* (pp. 131–138).

Lahav, O., & Mioduser, D. (2008a). Construction of cognitive maps of unknown spaces using a multi-sensory virtual environment for people who are blind. *Computers in Human Behavior*, *24*(3), 1139–1155.

Lahav, O., & Mioduser, D. (2008b). Haptic-feedback support for cognitive mapping of unknown spaces by people who are blind. *International Journal of Human-Computer Studies*, *66*(1), 23–35.

Lahav, O., Schloerb, D., Kumar, S., & Srinivasan, M. (2012). A virtual environment for people who are blind – a usability study. *Journal of Assistive Technologies*, *6*(1), 38–52.

Landau, B. (1986). Early map use as an unlearned ability. *Cognition*, *22*(3), 201–223.

Landau, B., Gleitman, H., & Spelke, E. (1981). Spatial knowledge and geometric representation in a child blind from birth. *Science*, *213*(4513), 1275–1278.

Landau, B., & Jackendoff, R. (1993). "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences*, *16*(2), 217 – 265.

Lawrence, M. M. (2011). *Behavioral and neurological studies in tactile map reading and training by persons who are blind or visually impaired.* Unpublished doctoral dissertation, University of Oregon.

Lawrence, M. M., Martinelli, N., & Nehmer, R. (2009). A haptic soundscape map of the university of oregon. *Journal of Maps*, *v2009*, 19–29.

Lederman, S. J., & Klatzky, R. L. (1987). Hand movements: A window into haptic object recognition. *Cognitive psychology*, *19*(3), 342–368.

Lederman, S. J., & Klatzky, R. L. (2009). Haptic perception: A tutorial. *Attention, Perception, & Psychophysics*, *71*(7), 1439–1459.

Lemon, O. (2011, April). Learning what to say and how to say it: Joint optimisation of spoken dialogue management and natural language generation. *Computer Speech & Language*, *25*(2), 210–221.

Leonard, J. A., & Newman, R. C. (1970). Three types of 'Maps' for blind travel. *Ergonomics*, *13*(2), 165–179.

Levelt, W. J. M. (1982). Linearization in describing spatial networks. In S. Peters & E. Saarinen (Eds.), *Processes, beliefs, and questions: Essays on formal semantics of natural language and natural language processing* (pp. 199–220). Doordrecht: D. Reidel.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: The MIT Press.

Levinson, S. C. (1996a). Frames of reference and Molyneux's question: Crosslinguistic evidence. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 109 – 169). Cambridge, MA: MIT Press.

Levinson, S. C. (1996b). Language and space. *Annual Review of Anthropology*, *25*(1), 353–382.

Liben, L. S. (1981). Spatial representation and behavior: Multiple perspectives. In L. S. Liben, A. H. Patterson, & N. Newcombe (Eds.), *Spatial representation and behavior across the life span: Theory and application* (pp. 3–36). New York: Academic Press.

Liben, L. S. (1991). Environmental cognition through direct and representational experiences: A life-span perspective. In G. W. Evans & T. Gärling (Eds.), *Environmenal cognition and action: An integrated approach* (pp. 245–276). Oxford: Oxford University Press.

Linsky, B. (2011). The notation in principia mathematica. In E. N. Zalta (Ed.), *The standford encyclopedia of philosophy* (Fall 2011 ed.). Retrieved from `http://plato.stanford.edu/archives/fall2011/entries/pm-notation/`

Lobben, A. K. (2004). Tasks, strategies, and cognitive processes associated with navigational map reading: A review perspective. *The Professional Geographer*, *56*(2), 270–281.

Lohmann, K. (2009). *Indexikalität und Evidenz visueller Zeichen. Das Beispiel der Fotografie im Übergang vom analogen zum digitalten Zeitalter.* Unpublished M.A. thesis, RWTH Aachen.

Lohmann, K. (2011). The use of sketch maps as measures for spatial knowledge. In J. Wang, K. Brölemann, M. Chipofya, A. Schwering, & J. O. Wallgrün (Eds.), *Understanding and processing sketch maps – proceedings of the COSIT 2011 workshop*

(Vol. 42, pp. 45–54). IOS Press.

Lohmann, K., Eichhorn, O., & Baumann, T. (2012). Generating situated assisting utterances to facilitate tactile-map exploration: A prototype system. In *Proceedings of the NAACL workshop on speech and language processing for assistive technologies 2012.* Montreal, QC.

Lohmann, K., Eschenbach, C., & Habel, C. (2011). Linking spatial haptic perception to linguistic representations: assisting utterances for tactile-map explorations. In M. Egenhofer, N. Giudice, R. Moratz, & M. Worboys (Eds.), *Spatial information theory* (pp. 328–349). Berlin, Heidelberg: Springer.

Lohmann, K., & Habel, C. (2012). Extended verbal assistance facilitates knowledge acquisition of virtual tactile maps. In C. Stachniss, K. Schill, & D. H. Uttal (Eds.), *Spatial cognition 2012* (pp. 299–318). Berlin, Heidelberg: Springer.

Lohmann, K., Kerzel, M., & Habel, C. (2010). Generating verbal assistance for tactile-map explorations. In I. van der Sluis, K. Bergmann, C. van Hooijdonk, & M. Theune (Eds.), *Proceedings of the 3rd workshop on multimodal output generation 2010.* Dublin, Ireland.

Lohmann, K., Kerzel, M., & Habel, C. (2012). Verbally assisted virtual-environment tactile maps: A prototype system. In C. Graf, N. A. Giudice, & F. Schmid (Eds.), *Proceedings of the workshop on spatial knowledge acquisition with limited information displays (SKALID 2012)* (pp. 25–30). CEUR-WS.org. Retrieved from `http://ceur-ws.org/Vol-888`

Lohmann, K., Yu, J., Kerzel, M., Wang, D., & Habel, C. (2013). Verbally assisting virtual-environment tactile maps: A cross-linguistic and cross-cultural study. In F. Sun, D. Hu, & H. Liu (Eds.), *Foundations and practical applications of cognitive systems and information processing. Advances in intelligent systems and computing (215).* Berlin, Heidelberg: Springer.

Loomis, J. M., Klatzky, R. L., Avraamides, M., Lippa, Y., & Golledge, R. G. (2007). Functional equivalence of spatial images produced by perception and spatial language. In F. Mast & L. Jäncke (Eds.), *Spatial processing in navigation, imagery, and perception* (pp. 29–48). New York, NY: Springer.

Loomis, J. M., Klatzky, R. L., & Giudice, N. A. (2012). Sensory substitution of vision: Importance of perceptual and cognitive processing. In R. Manduchi & S. Kurniawan (Eds.), *Assistive technology for blindness and low vision.* Boca Raton, FL: CRC Press.

Loomis, J. M., Klatzky, R. L., & Lederman, S. J. (1991). Similarity of tactual and visual picture recognition with limited field of view. *Perception*, *20*(2), 167–177.

Lötzsch, J. (1994). Computer-aided access to tactile graphics for the blind. In W. L. Zagler, G. Busby, & R. Wagner (Eds.), *Computers for handicapped persons: 4th international conference, ICCHP '94 (LNCS 860)* (pp. 575–581). Berlin, Heidelberg: Springer.

Lynch, K. (1960). *The image of the city.* Cambridge, MA; London: MIT Press.

MacEachren, A. M. (1992). Application of environmental learning theory to spatial knowledge acquisition from maps. *Annals of the Association of American Geogra-*

phers, *82*(2), 245–274.

MacEachren, A. M. (2004). *How maps work: representation, visualization, and design.* New York, NY: Guilford Press.

Magnusson, C., & Rassmus-Gröhn, K. (2004). A dynamic haptic-audio traffic environment. In *Proceedings of Eurohaptics 2004.* München, Germany.

Magnusson, C., & Rassmus-Gröhn, K. (2005). A virtual traffic environment for people with visual impairment. *Visual Impairment Research*, *7*(1), 1–12.

Mann, W. C. (1984). Discourse structures for text generation. In *Proceedings of the 10th international conference on computational linguistics* (pp. 367–375).

Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, *8*(3), 243–281.

Massie, T. H. (1993). *Design of a three degree of freedom force-reflecting haptic interface.* Massachusetts Institute of Technology.

Massie, T. H., & Salisbury, J. K. (1994). The PHANTOM haptic interface: A device for probing virtual objects. In *Proceedings of the ASME winter annual meeting, symposium on haptic interfaces for virtual environment and teleoperator systems.* Chicago, IL, USA.

May, M. (1992). *Mentale Modelle von Städten.* Münster: Waxmann.

McKeown, K. R. (1985). *Text generation: Using discourse strategies and focus constraints to generate natural language text.* Cambridge: Cambridge University Press.

McNamara, T. (1999). Single-code vs. multi-code theories in cognition. In R. J. Sternberg (Ed.), *Nature of cognition* (pp. 113–135). Cambridge, MA: MIT Press.

Meilinger, T. (2005). Wayfinding with maps and verbal directions. In *Proceedings of the 27th annual conference of the cognitive science society (CogSci 2005)* (pp. 1473–1478). Stresa, Italy.

Mellish, C. (2000). Understanding shortcuts in NLG systems. In *Proceedings of impacts in natural language generation: NLG between technology and applications* (pp. 43–50). Dagstuhl, Germany.

Meteer, M. W. (1991). Bridging the generation gap between text planning and linguistic realization. *Computational Intelligence*, *7*(4), 296–304.

Miele, J. A., Landau, S., & Gilden, D. (2006). Talking TMAP: automated generation of audio-tactile maps using Smith-Kettlewell's TMAP software. *British Journal of Visual Impairment*, *24*(2), 93–100.

Millar, S. (1994). *Understanding and representing space: Theory and evidence from studies with blind and sighted children.* Oxford; New York, NY: Oxford University Press.

Millar, S. (1995). Understanding and representing spatial information. *British Journal of Visual Impairment*, *13*(1), 8–11.

Montello, D. R. (1991). The measurement of cognitive distance: Methods and construct validity. *Journal of Environmental Psychology*, *11*(2), 101–122.

Montello, D. R. (1993). Scale and multiple psychologies of space. In A. Frank & I. Campari (Eds.), *Spatial information theory a theoretical basis for GIS (LNCS*

716) (pp. 312–321). Berlin, Heidelberg: Springer.

Montello, D. R. (1998). A new framework for understanding the acquisition of spatial knowledge in large-scale environments. In M. J. Egenhofer & R. G. Golledge (Eds.), *Spatial and temporal reasoning in geographic information systems* (pp. 143–154). New York: Oxford University Press.

Morash, V., Connell Pensky, A. E., Alfaro, A. U., & McKerracher, A. (2012). A review of haptic spatial abilities in the blind. *Spatial Cognition & Computation*, *12*(2-3), 83–95.

Moustakas, K., Nikolakis, G., Kostopoulos, K., Tzovaras, D., & Strintzis, M. G. (2007). Haptic rendering of visual data for the visually impaired. *IEEE Multimedia*, *14*(1), 62–72.

Münzer, S., & Hölscher, C. (2011). Entwicklung und Validierung eines Fragebogens zu räumlichen Strategien. *Diagnostica*, *57*(3), 111–125.

Newcombe, N. (1985). Methods for the study of spatial cognition. In R. Cohen (Ed.), *The development of spatial cognition* (pp. 277–300). Hillsdale, NJ; London: Lawrence Erlbaum.

Newcombe, N., & Huttenlocher, J. (2003). *Making space: The development of spatial representation and reasoning.* Cambridge, MA: The MIT Press.

Noordzij, M. L., Zuidhoek, S., & Postma, A. (2006). The influence of visual experience on the ability to form spatial mental models based on route and survey descriptions. *Cognition*, *100*(2), 321–342.

Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, *15*(5), 625–632.

O’Malley, M. K., & Goldfarb, M. (2002). Comparison of human haptic size identification and discrimination performance in real and simulated environments. In *Proceedings of haptic interfaces for virtual environment and teleoperator systems (HAPTICS 2002)* (pp. 10–17).

O’Malley, M. K., & Uppermann, G. (2006). A study of perceptual performance in haptic virtual environments. *Journal of Robotics and Mechatronics*, *4*(18), 10–17.

Paladugu, D. A., Wang, Z., & Li, B. (2010). On presenting audio-tactile maps to visually impaired users for getting directions. In *Proceedings of the 28th international conference on human factors in computing systems* (pp. 3955–3960). Atlanta, GA, USA.

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 259–303). Hillsdale, NJ: Lawrence Erlbaum.

Paneels, S., & Roberts, J. C. (2010). Review of designs for haptic data visualization. *IEEE Transactions on Haptics*, *3*(2), 119–137.

Parente, P., & Bishop, G. (2003). BATS: the blind audio tactile mapping system. In *Proceedings of the ACM southeast regional conference.* Pittsburg, PA, USA.

Parkes, D. (1988). “NOMAD”: an audio-tactile tool for the acquisition, use and management of spatially distributed information by partially sighted and blind people. In *Proceedings of the 2nd international conference on maps and graphics for visually* *disabled people.* Nottingham, UK.

Parkes, D. (1994). Audio tactile systems for designing and learning complex environments as a vision impaired person: static and dynamic spatial information access. *Learning Environment Technology: Selected Papers from LETA*, *94*, 219–223.

Pasqualotto, A., & Proulx, M. J. (2012). The role of visual experience for the neural basis of spatial cognition. *Neuroscience & Biobehavioral Reviews*, *36*(4), 1179–1187.

Passini, R., & Proulx, G. (1988). Wayfinding without vision. *Environment and Behavior*, *20*(2), 227.

Passini, R., Proulx, G., & Rainville, C. (1990). The spatio-cognitive abilities of the visually impaired population. *Environment and Behavior*, *22*(1), 91.

Peirce, C. S. (1960). *Collected papers of Charles Sanders Peirce* (4th ed.). London: Oxford University Press.

Perkins, C., & Gardiner, A. (2003, December). Real world map reading strategies. *Cartographic Journal, The*, *40*(3), 265–268.

Petrie, H., Johnson, V., Strothotte, T., Raab, A., Fritz, S., & Michel, R. (1996). MoBIC: designing a travel aid for blind and elderly people. *Journal of Navigation*, *49*(01), 45–52.

Pick, H. L., & Lockman, J. J. (1981). From frames of reference to spatial representations. In L. S. Liben, A. H. Patterson, & N. Newcombe (Eds.), *Spatial representation and behavior across the life span: Theory and application* (pp. 39–61). New York: Academic Press.

Pitt, D. (2008). Mental representation. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy.* Retrieved from `http://plato.stanford.edu/archives/fall2008/entries/mental-representation/`

Poole, D., & Macworth, A. (2010). *Artificial intelligence. foundations of computational agents.* Cambridge, MA: Cambridge University Press.

Posner, R. (1985). Nonverbale Zeichen in öffentlicher Kommunikation. *Zeitschrift für Semiotik*, *7*(3), 235–271.

Quinlan, P., & Dyson, B. (2008). *Cognitive psychology.* Essex: Prentice Hall.

Reiter, E. (1994). Has a consensus NL generation architecture appeared, and is it psycholinguistically plausible. In *Proceedings of the seventh international workshop on natural language generation* (pp. 163–170). Kennebunkport, ME, USA.

Reiter, E., & Dale, R. (2000). *Building natural language generation systems.* Cambridge: Cambridge University Press.

Reiter, E., Sripada, S., & Robertson, R. (2003). Acquiring correct knowledge for natural language generation. *Journal of Artificial Intelligence Research*, *18*(1), 491–516.

Rice, M., Jacobson, R. D., Golledge, R. G., & Jones, D. (2005). Design considerations for haptic and auditory map interfaces. *Cartography and Geographic Information Science*, *32*(4), 381–391.

Rich, C., Sidner, C. L., & Lesh, N. (2001). Collagen: applying collaborative discourse theory to human-computer interaction. *AI Magazine*, *22*(4), 15–26.

Röder, B., & Rösler, F. (1998). Visual input does not facilitate the scanning of spatial images. *Journal of Mental Imagery*, *22*(3–4).

Röder, B., Rösler, F., Grunwald, M., & Beyer, L. (2001). Ein Vergleich haptischer Wahrnehmungsleistungen zwischen blinden und sehenden Personen. In *Der bewegte Sinn: Grundlagen und Anwendung der haptischen Wahrnehmung.* Berlin, Heidelberg: Springer.

Roy, D., & Reiter, E. (2005). Connecting language to the world. *Artificial Intelligence*, *167*(1-2), 1–12.

Russell, B. (1905). On denoting. *Mind*, *14*(56), 479–493.

Salisbury, K., Brock, D., Massie, T., Swarup, N., & Zilles, C. (1995). Haptic rendering: Programming touch interaction with virtual objects. In *Proceedings of the 1995 symposium on interactive 3D graphics* (pp. 123–130).

Scaife, M., & Rogers, Y. (1996). External cognition: how do graphical representations work? *International Journal of Human–Computer Studies*, *45*(2), 185–213.

Schmitz, B., & Ertl, T. (2012). Building augmented you-are-here maps through collaborative annotations for the visually impaired. In C. Graf, N. A. Giudice, & F. Schmid (Eds.), *Proceedings of the workshop on spatial knowledge acquisition with limited information displays (SKALID 2012)* (pp. 13–18). CEUR-WS.org. Retrieved from `http://ceur-ws.org/Vol-888`

Schneider, J. G., & Stöckl, H. (2011). Medientheorien und Multimodalität: Zur Einführung. In J. G. Schneider & H. Stöckl (Eds.), *Medientheorien und Multimodalität. Ein TV-Werbespot - Sieben methodische Beschreibungsansätze.* Köln: Halem.

Schröder, M., & Trouvain, J. (2003). The German text-to-speech synthesis system MARY: a tool for research, development and teaching. *International Journal of Speech Technology*, *6*(4), 365–377.

Shelton, A. L., & McNamara, T. P. (2001). Systems of spatial reference in human memory. *Cognitive Psychology*, *43*(4), 274–310.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psycholocial Bulletin*, *86*(2), 420–428.

Siegel, A. W., & White, S. H. (1975). The development of spatial representations of large-scale environments. *Advances in child development and behavior*, *10*, 9–55.

Siekierska, E., McCurdy, W., & Peterson, M. P. (2008). Internet-based mapping for the blind and people with visual impairment. In M. P. Peterson (Ed.), *International perspectives on maps and the internet* (pp. 283–300). Berlin, Heidelberg: Springer.

Siekierska, E., & Müller, A. (2003). Tactile and audio-tactile maps within the canadian 'Government on-line' program. *The Cartographic Journal*, *40*(3), 299–304.

Sjöström, C. (2001a). Designing haptic computer interfaces for blind people. In *Sixth international symposium on signal processing and its applications.* Kuala Lumpur, Malaysia.

Sjöström, C. (2001b). Using haptics in computer interfaces for blind people. In *CHI'01 extended abstracts on human factors in computing systems* (pp. 245–246). Seattle, WA, USA.

Stamm, M., Altinsoy, M., & Merchel, S. (2011). Influence of the auditory localization direction on the haptic estimation of virtual length. *Haptic and Audio Interaction*

Design, 101–109.

Stetter, C. (2005). *System und Perfomanz.* Weilerswist: Velbrück Wissenschaft.

Stevens, A., & Coupe, P. (1978). Distortions in judged spatial relations. *Cognitive psychology*, *10*(4), 422–437.

Striegnitz, K., Denis, A., Gargett, A., Garoufi, K., Koller, A., & Theune, M. (2011). Report on the second second challenge on generating instructions in virtual environments (GIVE-2.5). In *Proceedings of the 13th European workshop on natural language generation* (pp. 270–279). Nancy, France.

Struiksma, M. E. (2011). *On the language of space. Neurocognitive studies in blind and sighted individuals.* Unpublished doctoral dissertation, Univseriteit Utrecht.

Talmy, L. (2000). How language structures space. In *Toward a cognitive semantics. Vol. 1: Concept structuring systems* (pp. 177–254). Cambridge, MA, USA: MIT Press.

Tappe, H., & Habel, C. (1998). Verbalization of dynamic sketch maps: Layers of representation and their interaction. In *Proceedings of the 20th annual meeting of the cognitive science society (CogSci-98).* Madison, WI, USA.

Tatham, A. F. (2003). Tactile mapping: Yesterday, today and tomorrow. *The Cartographic Journal*, *40*(3), 255–258.

Taylor, H. A., & Tversky, B. (1992a). Descriptions and depictions of environments. *Memory and Cognition*, *20*(5), 483–496.

Taylor, H. A., & Tversky, B. (1992b). Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language*, *31*(2), 261–292.

Taylor, H. A., & Tversky, B. (1996). Perspective in spatial descriptions. *Journal of Memory and Language*, *35*(3), 371–391.

Tenbrink, T., Bergmann, E., & Konieczny, L. (2011). Wayfinding and description strategies in an unfamiliar complex building. In *Proceedings of the the 33rd annual conference of the cognitive society* (p. 1262-1267). Boston, MA, USA.

Tenbrink, T., & Winter, S. (2009). Variable granularity in route directions. *Spatial Cognition & Computation*, *9*(1), 64–93.

Thorndyke, P. W., & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. *Cognitive psychology*, *14*(4), 560–589.

Thorndyke, P. W., & Stasz, C. (1980). Individual differences in procedures for knowledge acquisition from maps. *Cognitive Psychology*, *12*(1), 137–175.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological review*, *55*(4), 189–208.

Tschander, L. B., Schmidtke, H. R., Eschenbach, C., Habel, C., & Kulik, L. (2003). A geometric agent following route instructions. In C. Freksa, W. Brauer, C. Habel, & K. Wender (Eds.), *Spatial cognition III.* Lecture Notes in Computer Science, Vol. 2685 (pp. 90–111). Berlin, Heidelberg: Springer.

Tversky, B. (1992). Distortions in cognitive maps. *Geoforum*, *23*(2), 131–138.

Tversky, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. In A. U. Frank & I. Campari (Eds.), *Spatial information theory: A theoretical basis for GIS* (pp. 14–24). Berlin: Springer.

Ungar, S. (2000). Cognitive mapping without visual experience. In R. Kitchin & S. Freundschuh (Eds.), *Cognitive mapping: Past, present and future* (pp. 221–248). London: Routledge.

Ungar, S., Blades, M., & Spencer, C. (1995). Visually impaired children's strategies for memorising a map. *British Journal of Visual Impairment*, *13*(1), 27–32.

Ungar, S., Blades, M., & Spencer, C. (1997). Strategies for knowledge acquisition from cartographic maps by blind and visually impaired adults. *The Cartographic Journal*, *34*(2), 93–110.

van Deemter, K., Krahmer, E., & Theune, M. (2005). Real versus template-based natural language generation: a false opposition? *Computational Linguistics*, *31*(1), 24.

Wang, Z., Li, B., Hedgpeth, T., & Haven, T. (2009). Instant tactile-audio map: enabling access to digital maps for people with visual impairment. In *Proceedings of the 11th international ACM SIGACCESS conference on computers and accessibility* (pp. 43–50). Pittsburg, PA, USA.

Wellek, S. (2003). *Testing statistical hypotheses of equivalence.* Boca Raton; London; New York; Washington, D.C.: CRC Press.

Wen, W., Ishikawa, T., & Sato, T. (2011). Working memory in spatial knowledge acquisition: Differences in encoding processes and sense of direction. *Applied Cognitive Psychology*, *25*(4), 654–662.

Werner, S., Krieg-Brückner, B., Mallot, H. A., Schweizer, K., & Freksa, C. (1997). Spatial cognition: The role of landmark, route, and survey knowledge in human and robot navigation. In *GI Jahrestagung 1997* (pp. 41–50). Berlin: Springer.

Zeng, L., & Weber, G. (2010). Audio-haptic browser for a geographical information system. In K. Miesenberger, J. Klaus, W. Zagler, & A. Karschmer (Eds.), *Computers helping people with special needs, part II* (pp. 466–473). Berlin, Heidelberg: Springer.

Zeng, L., & Weber, G. (2012). Building augmented you-are-here maps through collaborative annotations for the visually impaired. In C. Graf, N. A. Giudice, & F. Schmid (Eds.), *Proceedings of the workshop on spatial knowledge acquisition with limited information displays (SKALID 2012)* (pp. 7–12). CEUR-WS.org. Retrieved from `http://ceur-ws.org/Vol-888`