TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Steuerungs- und Regelungstechnik

Extraction of Probabilistic Route Information Representations from Human-Robot Dialogs

Andrea M. Bauer

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing. habil. Gerhard Rigoll

Prüfer der Dissertation:

1. TUM Junior Fellow Dr.-Ing. Dirk Wollherr
2. Univ.-Prof. Dr. Manfred Tscheligi
   Universität Salzburg / Österreich

Die Dissertation wurde am 21.04.2010 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 26.11.2010 angenommen.

Foreword

This thesis is the result of three years of research at the Institute of Automatic Control Engineering (LSR) of Technische Universität München. I am grateful to all people who contributed to my research and supported me during this time.

First of all, I would like to thank my doctoral advisor Dr. Dirk Wollherr for all open discussions and constructive feedback. I would like to express my gratitude towards Prof. Martin Buss who trusted in my abilities and provided an inspiring working environment for my research. For the great interdisciplinary collaboration I thank Dr. Astrid Weiss and Prof. Manfred Tscheligi of the ICT&S Center, University of Salzburg. For their advice on experimental design and statistics I would like to thank my colleague Raphaela Groten and Stephan Haug from TUMStat. Great thanks go to Prof. Niall Palfreyman for the thorough proofreading of this thesis.

I enjoyed the pleasant and supportive working environment at the LSR to which all of my colleagues have contributed. For the great team work, fruitful discussions, and assistance, I would like to thank the other members of the *ACE*-team: Klass Klasing, Georgios Lidoris, Quirin Mühlbauer, Florian Rohrmüller, Stefan Sosnowski, Tingting Xu, and Tianguang Zhang. Special thanks go to my colleagues Michelle Karg, Markus Rank, Michael Scheint, Ulrich Unterhinninghofen, and Tingting Xu, who have supported me with their technical knowledge, proofreading of parts of this thesis, encouragement, and most importantly friendship. I would like to thank all students who contributed to this thesis: Adrian Garcea, Barbara Gonsior, Raul Heinrich, Lilian Kettner, Taru Laamannen, and Jens-Nikolas Martens.

For always supporting me, I would like to thank my mom who is my technical hero. And last but not least, I thank my boyfriend Philipp for his love and encouragement.

Munich, April 2010 Andrea Bauer

To my grandmother Marie Fetterová
and my dog Hexi.

Contents

Abstract

This thesis addresses methods that enable robots to extract missing route knowledge by asking humans for directions. In a first step an experiment was conducted in which a robot had to navigate to a designated goal in an unknown urban environment without any prior map knowledge or GPS, but solely by asking passers-by for directions. While the experiment was successful, at the same time the results point up further challenges for extracting route information through human-robot communication. Hence, research questions are derived that are answered in the remainder of the thesis, namely proactive extraction of route information from human-robot dialogs, probabilistic representation of route information, and reasoning about extracted route descriptions. Linguistic principles relevant to direction-inquiry dialogs are reviewed. Guidelines for human-robot dialogs are derived from these and implemented in a dialog system. This renders the dialog natural to humans and facilitates proactive extraction of unambiguous route information. As route information is usually simplified and distorted, probabilistic models of individual route information are developed. Certainty values are computed from direction and error probabilities to assess the reliability of direction information. Quantitative and qualitative distance information is modeled by posterior probability distributions that assess the accuracy of distance information. Finally, a system is developed that allows robots to reason about different route descriptions. Route descriptions are represented as graphs and evaluated for plausibility by pattern matching and route similarity assessment. Reasonable route information is included in the route belief of the robot, while conflicting information is inquired about if necessary. All methods are evaluated experimentally with collected data or by human participants. The methods and systems introduced are designed in adaptable ways and can be expanded for extracting and representing other types of information from human-robot dialogs.

Zusammenfassung

Diese Arbeit beschäftigt sich mit Methoden, die es Robotern ermöglichen, sich fehlendes Wegwissen anzueignen indem sie Menschen gezielt nach dem Weg fragen. Zunächst wurde ein Experiment durchgeführt, in dem ein Roboter seinen Weg zu einem vorgegebenen Ziel in in einer unbekannten Umgebung ohne vorheriges Kartenwissen oder GPS, sondern allein durch Wegangaben von Passanten erreichen musste. Dieses Experiment war erfolgreich, zeigt aber auch auf, welche Herausforderungen für die Aneignung von Wegwissen durch Mensch-Roboter-Kommunikation noch bestehen. Daraus leiten sich die Forschungsschwerpunkte dieser Arbeit ab, nämlich Extraktion von Weginformationen aus Mensch-Roboter-Dialogen, probabilistische Repräsentation von Weginformationen und das Schlussfolgern über gegebene Wegbeschreibungen. Aus linguistischen Erkenntnissen werden Richtlinien für Mensch-Roboter-Dialoge über Wegwissen abgeleitet und in einem Dialogsystem implementiert. Dies ermöglicht eine für den Menschen natürliche Kommunikation und die zuverlässige Extraktion von inhaltlich eindeutigen Weginformationen. Da Weginformationen üblicherweise Schätzwerte und Vereinfachungen der tatsächlichen Werte wiedergeben, werden in einem nächsten Schritt probabilistische Modelle für einzelne Weginformationen erstellt. Aus Richtungs- und Fehlerwahrscheinlichkeiten werden Zuverlässigkeitswerte für

Richtungsangaben berechnet. Quantitative und qualitative Distanzangaben werden durch bedingte Wahrscheinlichkeitsdichteverteilungen modelliert, die Aufschluss über die Genauigkeit der einzelnen Angaben liefern. Zuletzt wird ein System vorgestellt, das es Robotern ermöglicht, verschiedene Wegauskünfte zu vergleichen. Dabei werden die Wegauskünfte graphentheoretisch repräsentiert und über Pattern-Matching und Wegvektorähnlichkeit auf Plausibilität untersucht. Sinnvolle Weggraphen werden in die Wissensbasis aufgenommen, während das System gegebenenfalls gezielt bei widersprüchlichen Informationen nachfragt. Alle Methoden werden mit durch Umfragen erhobene Daten oder mit Probanden experimentell evaluiert. Die Vorgestellten Systeme und Verfahren sind modular gehalten und lassen sich auf Aneignung und Repräsentation allgemeiner Informationen durch Mensch-Roboter-Dialoge erweitern.

Notations

Abbreviations

ACE	*Autonomous City Explorer*
ANOVA	analysis of variance
CONCL	dialog system state: *conclusion*
CONF	dialog system state: *confirmation*
FSM	finite state machine
GIVDIR	dialog system state: *give directions*
GPS	Global Positioning System
GUI	graphical user interface
DG	dialog guideline
HRI	human-robot interaction
INTRO	dialog system state: *introduction*
pdf	probability density function
SRAM	*Simultaneous Reasoning and Mapping*
s.t.	subject to
SSH	Spatial Semantic Hierarchy
XML	Extensible Markup Language

Conventions

Scalars, Vectors, and Matrices

Scalars are denoted by letters in italic type. *Vectors* are denoted by bold lower case letters in italic type, vector $\boldsymbol{x}$ is composed of elements x_i. *Matrices* are denoted by bold upper case letters in italic type, matrix $\boldsymbol{M}$ is composed of elements M_{ij} (i^{th} row, j^{th} column).

x, X	scalar
$\boldsymbol{x}$	vector
$\boldsymbol{X}$	matrix
$\boldsymbol{X}^T$	transposed of $\boldsymbol{X}$
$\boldsymbol{x} \circ \boldsymbol{y}$	Hadamard product of $\boldsymbol{x}$ and $\boldsymbol{y}$
$g(\cdot)$	scalar function
$\lvert\cdot\rvert$	absolute value
$\lVert\cdot\rVert$	Euclidean norm

List of Symbols

Roman Symbols

A_i	path in B
$\boldsymbol{A}_i$	vector spanned by path A_i
$\boldsymbol{a}_{i,j}$	edge in path A_i as vector
B	belief
B_H, B_R	belief of human/robot
$B_{\mathrm{prel},i}$	preliminary belief
b_c^h	border between *'here'* and *'close'*
b_f^c	border between *'close'* and *'far'*
C^B, C^R	conflicting information in B, R
c_k	certainty value
c_k^M	certainty value in a metric graph
c_k^{sensor}	confidence value of sensor data
$D_M(x, y)$	pattern matching metric
d_k	length of edge
d_k^M	length of edge in metric graph
d_{est}	estimated distance
d_{real}	real distance
E_k	edge
E_k^M	edge in metric graph
E_k^T	edge in topological graph
$e(M_{i,j})$	number of edges in $M_{i,j}$
e_d^A	relative absolute error of distance estimate
e_t^A	relative absolute error of time estimate
e_d^C	relative constant error of distance estimate
e_t^C	relative constant error of time estimate
f_c	frequency of qualitative distance *'close'*
$F_{E,k}$	cumulative error frequency
$f_{E,k}$	error frequency
f_f	frequency of qualitative distance *'far'*
f_h	frequency of qualitative distance *'here'*
G	route graph
g_{rat}	rational function
g_{pow}	power function
H_i	human participant
k	route segment
l_i	landmark type
l_i^M	landmark type in metric graph
$M_{i,j}$	sub-match
N_i	node
O_H	personal reference system of human

O_{HR}	common reference system of human and robot
O_R	personal reference system of robot
$\boldsymbol{P}$	transition probability matrix
P, p	probability
$P_{D,k}$	direction probability
$P_{D,\mathrm{rel}}$	relative direction probability
$P_{E,k}$	error probability
$P(d)$	marginal probability of distance information d
$P(d\|d_{\mathrm{real}})$	conditional probability of distance information d
$P(d_{\mathrm{real}})$	prior probability of distance information d
$P(d_{\mathrm{real}}\|d)$	posterior probability of distance information d
$P_{\mathrm{rel}}(d_{\mathrm{real}}\|d)$	relative posterior probability
p_{ij}	transition probability between states $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$
R	route
$\boldsymbol{R}$	vector spanned by route graph R
$\boldsymbol{r}_k$	edge in path R as vector
$r_{y_{\mathrm{est}}}^{\mathrm{pow}}$	sum of squared residuals of power function
$r_{y_{\mathrm{est}}}^{\mathrm{rat}}$	sum of squared residuals of rational function
$\boldsymbol{r}_k$	edge in route r as vector
r_{real}	range of real distances
S	set of direction states
$S(\boldsymbol{m}, \boldsymbol{n})$	similarity function/metric
$S_D(\boldsymbol{m}, \boldsymbol{n})$	direction similarity metric
$S_M(\boldsymbol{m}, \boldsymbol{n})$	magnitude similarity function
$S_V(\boldsymbol{m}, \boldsymbol{n})$	vector similarity function
$\boldsymbol{S}_V$	vector similarity matrix
$\boldsymbol{s}_i$	direction state
${}^{H}T_{HR}$	coordinate transformation from O_H to O_{HR}
${}^{R}T_{HR}$	coordinate transformation from O_R to O_{HR}
t_{est}	estimated time
$\boldsymbol{u}$	random effects
$\boldsymbol{v}_k$	global direction state vector
$\tilde{\boldsymbol{v}}_k$	direction prediction based on overall previous route
v_{walk}	walking velocity
$\boldsymbol{w}$	weight vector
w_D	weight for direction probability
w_E	weight for error probability
$\boldsymbol{X}$, $\boldsymbol{Z}$	design matrices
$\boldsymbol{x}_k$	direction state vector
x_i	fraction of discrete direction
$\hat{\boldsymbol{x}}_k$	prediction of direction state at segment k
$\tilde{\boldsymbol{x}}_k$	direction prediction based on previous route segment
x_R, y_R	coordinates in the reference system of a robot
$\boldsymbol{y}$	data

Greek Symbols

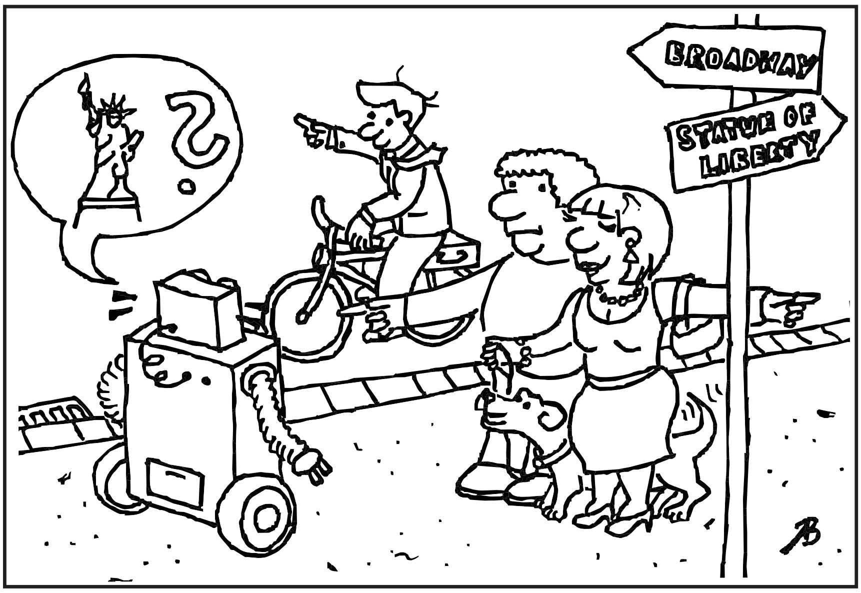
α_1, α_2	variables of g_{rat}
α_f, β_f	variables of $P(\textit{'far'}\vert d_{\text{real}})$
$\boldsymbol{\beta}$	effects of variables
β_1, β_2	variables of g_{pow}
$\gamma\,(M_{i,j})$	diagonal of match $M_{i,j}$
δ_k	direction of edge/vector
δ_k^M	direction in metric graph
$\boldsymbol{\epsilon}$	random errors
μ	mean value
μ_d, σ_d	variables of $P(d_{\text{est}}\vert d_{\text{real}})$
μ_E, σ_E	variables of $P_{E,k}$
μ_h	variable of $P(\textit{'here'}\vert d_{\text{real}})$
μ_c, σ_c	variables of $P(\textit{'close'}\vert d_{\text{real}})$
μ_t, σ_t	variables of $P(t_{\text{est}}\vert d_{\text{real}})$
$\boldsymbol{\pi}$	initial probability vector
π_i	initial state probability
σ	standard deviation
τ_e	threshold for the number of edges in a sub-match
τ_M	pattern matching threshold
$\tau_{P_{\text{rel}}}$	threshold for relative posterior
τ_S	threshold for highest similarity value
τ_{Sv}	threshold for similarity value

List of Figures

List of Tables

1 Introduction

Asking humans for information is a reasonable approach for robots to come by missing knowledge about complex, dynamic human-populated environments. This chapter provides an introduction to the thesis, reviews the state of the art of related research fields, and discusses the main scientific contributions.



1.1 Motivation and Challenges

It is a longstanding research goal to develop service robots that aid human users in their everyday lives [1, 2, 6]. Tasks that humans are envisioned to be relieved of by robots are daily chores such as preparing meals, loading dishwashers, and shop for groceries. However, service robots assisting humans in such intelligent and versatile ways may still be decades away. One of the reasons for this is that service robots need to be able to operate in human-populated environments which are unstructured, complex, and highly dynamic and a huge amount of up-to-date information is needed by robots operating in such settings.

Standard approaches for robots to come by knowledge, necessary to fulfil a task, are pre-programming and machine learning. As not every bit of information can be contained in program code, pre-programming is not feasible in real environments. Since complex dynamic environments would require a vast number of training sets to be able to make intelligent decisions, machine learning is not efficient in such scenarios. Therefore these standard approaches are not suitable to provide robots with all information necessary to operate in real-world environments. An alternative for robots to close gaps in their task knowledge is to ask humans for missing information, as this is a fast way to come by relevant and up-to-date information. Task descriptions extracted from human-robot dialogs can then be translated into machine understandable instructions, i.e. broken down into a sequence of action instructions. This approach is relatively new in robotics, as opposed to the more common unilateral scenario of robots solely aiding humans and providing information to them. Therefore there are still many unsolved problems and challenges. In particular these challenges include proactively extracting missing information from often vague and ambiguous human-robot communication, representing the information given by humans in a way that can be processed by technical systems, and reasoning about it in order to identifying inaccurate or erroneous information. Robots that overcome these problems are able to act more robustly in the face of new and unforeseen tasks and situations, and thus reach higher levels of autonomy and dependability.

This thesis addresses the challenges posed by the general problem of extracting task information from human-robot dialogs, based on the special case of extracting route information for navigation tasks. This is a prerequisite to reaching a goal location in an unknown or changing environment, as it provides robots with a basis for autonomous navigation and global path planning. Even with access to the internet or GPS systems, robots are bound to have gaps in their map knowledge of complex and dynamic human-populated environments. Applications that can benefit from this are for example running errands and shopping. Such applications are of use to personal service robotics and elderly care, where human staff can be relieved of such tasks and gain time that could be used for medical and psychological treatment as well as to keep the persons in their care company.

This thesis focuses on extracting a probabilistic route information representation from human-robot dialogs. In particular the presented research includes a dialog system for robots asking for directions, probabilistic models for individual route information, and a system for reasoning about route descriptions and simultaneously building a route belief as an internal representation of the plausible route descriptions. The dialog system includes findings from human-human communication, in order to render the dialogs more natural to humans and to unambiguously extract route information. The probabilistic models provide robots with means to assess individual directions and distances for reliability and accuracy. Finally, a system for reasoning about route information compares different route descriptions, assesses them for plausibility, inquires about conflicting information if necessary, and builds a route belief as a representation of the assumed route within the environment.

In this chapter the research areas related to the thesis are reviewed. This chapter does not provide a comprehensive overview of the state of the art, as the more specific research developments relevant to each chapter are covered at the beginning of each respective chapter. Overviews of the broad research fields of human-robot interaction and spatial cognition and computation are provided in the following. A comprehensive view of human-robot interaction (HRI), with design implications, ways of communication, and applications, is necessary for providing a robot with the ability to ask humans for directions, as these factors can affect the outcome of human-robot dialogs. An overview of spatial cognition and computation is needed to understand the ways that humans and technical systems represent and process spatial information.

This chapter is structured as follows. An overview of the state of the art in human-robot interaction is provided in Section 1.2. The state of the art in spatial cognition and computation is reviewed in Section 1.3. Finally, Section 1.4 presents the main contributions and the outline of the thesis.

1.2 Overview of Human-Robot Interaction

Human-robot interaction is a wide research field and constitutes a framework of insights useful to proactively extract information from human-robot dialogs. HRI includes interdisciplinary cooperation between classical robotics, cognitive sciences, and psychology and is a field with many implications for design and communication and with various applications. As the research done in this field is vast, only the aspects most relevant to this thesis are surveyed here. Extensive overviews of interactive robots and HRI are provided for example by Fong et al. [45, 178] and Kiesler et al. [79].

The first impression a human gets about a robot is how it is designed, i.e. the appearance of the robot, whether social distances are respected, and the subjective feeling of safety.

1.2.1 Design Implications

Design specifications for interactive robots are manifold and different from purely functional industrial robots. Among the issues that have to be considered when designing robots for HRI are the safety of humans, social distances, and robot appearance.

Safety is a crucial aspect of the design of interactive robots [97, 186]. A robot must not endanger a human under any circumstances. Since the very beginning of robotics this has been an evident fact and was formulated by Isaac Asimov in his famous *Three Laws of Robotics* [7]. On the one hand, the hardware of the robot must not be dangerous for humans, banning sharp edges, points and accessible electric current. On the other hand, control theoretic methods and suitable software design have to be applied to guarantee safety even in the face of disturbances or unforeseen events [96].

Robots interacting with human partners, do not only have to respect safety distances, but also social distances. Such distances are studied in proxemics [63] for interaction between humans. Proxemics discern between different social distances [64], i.e. intimate, personal, social, and public distance. Proxemic studies are performed with humans interacting with robots [177, 178] in order to define and control interaction distances that feel comfortable to humans and thus in theory improve the human's subjective feeling and the performance of the interaction.

Thought has to be given to the appearance of interactive robots, as it might effect the expectation of the human in the robot, and therefore on the long run the outcome of the human-robot interaction. Robot appearance design is studied for example from a traditional design theoretic view [176], for humanoid robots [179], and for robot heads [38]. Depending on the purpose and character of the interaction, robots are designed to look more or less similar to humans. Mori [124] found a function, called the "uncanny valley" that maps similarity to acceptance. He states that acceptance increases the more similar a robot is to a human, until a point is reached where the similarity is not quite 100 % and subtle deviations from human appearance and behaviour create an unnerving effect.

Contrary to design policies that attempt to make robots more similar to humans a study by Yamada et al. [204] suggest that robots with typical robot appearance and a primitive expression of emotions are more easily understood by humans than more life-like appearing robots with complex ways of expressing emotions, concluding that the appearance and expressions of robots should resemble their abilities. Another aspect that has to be taken into account, when designing robots, is the culture a robot is supposed to work in, as there are differences in the acceptance of robots by people with different cultural backgrounds [77]. Especially in Japan robots are accepted more positively than in Europe and America.

Another important aspect that has to be taken into account when designing robots for HRI is communication, as it allows for exchanging information during interaction. Human-robot communication is surveyed below.

1.2.2 Human-Robot Communication

Communication is a crucial aspect of interactive robots, since efficient interaction necessitates the exchange of complex information. Communication can be distinguished by the *specificity*, i.e. implicit or explicit information; the *direction* of the information exchange, i.e. the roles of human and robots as emitter and receiver of information; and the *mode* of communication, the main modes being haptic signals, physiological signals, gestures, and natural language.

Haptic communication can be maintained when the partners are connected physically either directly or through an object they both manipulate. Haptic communication usually conveys explicit information. It occurs through applied forces and torques or joint angles and orientations. It is studied from different perspectives, such as joint object manipulation [87, 88], dancing [53, 86] or handshaking [99].

Implicit information such as intentions and emotions are also communicated involuntarily through physiological signals such as heart rate and brain or muscle activity. Physiological signals can be analyzed by robots to deduce the level of approval and arousal [95] or the stress level [146]. Based on this information, robots can select appropriate actions.

Gestures are used to convey information implicitly or explicitly. Explicit gestures hold complex information, and are communicative gestures such as pointing gestures, primitive signs, or sign language. Implicit gestures include manipulative gestures and facial expressions. Robots can apply gesture recognition to extract information from gestures. Pointing gestures can be interpreted for example from camera images by matching an arm model [137] or interpreting kinematic information [113]. Some HRI approaches use a set of predefined gestures which the robot is able to recognize [84]. Intentions of humans can be inferred by robots from manipulative gestures [105], from head gestures such as eye gaze direction [156], or from facial expressions [76]. Some robots are able to use facial expressions [17, 169] for communication in return and thereby act socially. There are also robots that can generate hand and arm gestures [65] to communicate with humans.

Natural language communication provides explicit information through words and sentences, but can also provide information about emotions implicitly. Methods for processing and generating natural language are reviewed in detail below.

Natural Language Communication

For humans a very natural way to interact and exchange information is through speech, as it is suitable to efficiently convey complicated information. Therefore to render human-robot communication efficient and natural to humans, interactive robots need to be able to communicate through natural language. Natural language communication can convey information both explicitly and implicitly [208]. Implicit information in speech signals can be extracted by robots through emotion recognition [207] and also encoded in synthesized speech [161]. In the following, techniques and methods for explicit natural language communication are presented, as most of the relevant information is conveyed over this channel. A detailed overview of the history of explicit natural-language communication with computers is given, e.g. in [140]. Speech and language processing is reviewed in [75].

The research field of computational linguistics studies language from a computational perspective, thus connecting linguistics and computer engineering. Extensive overviews of computational linguistics are given for example in [13, 33]. Natural language processing embraces many major subfields of computational linguistics, such as speech recognition, text pre-processing, machine translation, information extraction, text generation, and speech synthesis. Speech recognition is concerned with the conversion of an acoustic signal captured by a microphone to a set of words. An overview of the developments in speech recognition technologies is presented in [73, 131]. Speech synthesis allows technical systems to produce speech; a broad overview of speech synthesis methods is presented in [103]. Text generation is a relatively new subfield of natural language processing, and deals with generating texts from symbolic formal representations; an overview of the main methods is provided in [9]. The field of text pre-processing includes techniques such as text parsing, spell checking, and referencing. Machine translation is concerned with translating texts from one language to another; an overview of techniques and tools is given in [69]. Speech synthesis and speech recognition is surveyed in [131].

Speech understanding is regarded as the major problem of natural language processing as it is the most general and complex task. Speech understanding requires a great variety of knowledge about the environment, context, speaker, topic, lexical frequency, semantically related topics, to name just a few aspects. Important approaches to speech understanding are semantic networks [202] that represent semantics as nodes connected by directed edges, and conceptual dependency [160] that is based on semantic networks, but uses edges that represent different dependencies. Recent approaches use the availability of high-performance computing systems with large spoken and written data collections, e.g. [114, 145], and by using statistical machine learning techniques [138, 190].

A sub-field of speech understanding is information extraction. Information extraction deals with the automatic extraction of structured information from unstructured machine readable text [159]. It is a special case of information retrieval [166] which is the search for content in large document databases. Information extraction techniques are surveyed in [159]. It has recently expanded into the domain of human-robot interaction, e.g. [74].

Human-robot communication in general and information extraction in specific are key requirements for robots that depend on task information from humans to achieve their goals. These abilities are useful in a wide variety of applications of HRI. The main applications are reviewed below.

1.2.3 Applications

Applications for HRI are manifold, with the main application areas in home service, healthcare, tour guiding, entertainment, construction, as well as search and rescue. These areas have different requirements and the degree to which they are presently integrated varies. Robots with any of these applications benefit from the ability to extract and represent missing task information, as it provides robots with a greater autonomy and a robustness against gaps in the task knowledge.

Home service will be a major application area for HRI in the future. First commercially available robots have been developed that vacuum-clean [153] and mow the lawn [150]. The household robot Wakamaru [195] developed by Mitsubishi is commercially available as well. It can connect itself to the home network to provide information for daily life, look after the house while the inhabitants are absent, and communicate with humans. A robot that acts as home security [206] is being developed. Truly collaborative home service robots are still a matter of research, such as robots assisting humans in the kitchen [11, 135].

Diverse healthcare robots are being developed, such as robots that guide the blind [70, 98, 189], robotic walkers [152], wheelchairs [71], elderly care robots [121, 142, 193], and robots for the therapy of autistic children [200]. These robots are all designated to aid diseased or fragile humans and must be especially well designed and scrutinized. Ethical issues have to be considered particularly in the field of healthcare and medicine.

Some robots are used as tour guides in Museums, i.e. RHINO [24] and its successor MINERVA [181], and office environments [116, 134]. They can guide humans to certain rooms or exhibitions and provide information. Toomas [59] is a robot guiding humans through a home improvement store. Tour guide robots need to be able to autonomously navigate through constrained human populated environments, and to communicate with humans in natural and intuitive ways.

Various entertainment robots capable of HRI are already commercially available, such as robotic pets [143, 170] and humanoid robots [100, 151].

A seminal field for HRI is construction, where robots can support human construction workers in carrying heavy loads and ease repetitive construction tasks. First steps toward joint construction have been taken by the mobile robot helper [88] that can share loads and handle objects in collaboration with a human. The JAST system [48] is capable of solving a small scale construction task collaboratively with a human.

Another field in which robots support humans and take on dangerous tasks is urban search and rescue [128, 129]. Robots are designed to move into collapsed buildings, collect data, and try and find human victims, who can then be rescued by human staff members.

Though most of these systems are still under development, it is only a matter of time until truly assistive and interactive robots will be commercially available and part of our everyday lives. Such robots need to be able to assisting humans in intelligent and versatile ways [19]. For this purpose robots need the ability to extract missing information from HRI. Asking humans for information about their spatial environments will enable robots to operate in unknown environments and fulfill new and unforeseen tasks.

1.2.4 Interactive Robots Extracting Spatial Information

Recently researchers have started developing robots that are able to extract information about their spatial environments from HRI. A robot that creates a map of its environment by exploration and asks humans to label areas of interest is presented in [68]. The robot Biron [171] integrates spoken dialog and visual localization to learn and label new places. Both robots use HRI as a means to attach human understandable labels to their sensor-based spatial belief. Also there are robots that ask humans for directions in structured indoor environments. A wheelchair robot [119] can be given coarse qualitative route descriptions. The office robot Jijo-2 [8] learns the locations of offices and staff by moving around and asking humans for information. A robot asking for the way at a robotics conference is presented in [118], and was the winner of the AAAI-2002 robot challenge [94]. A global inference approach [197] aims at having a robot automatically find a path within an office environment based upon human directions. Finally, a miniature robot that can find its way in a model town by asking humans for directions is described in [101]. All of these robots operate in highly structured and often static indoor environments, and only a few use natural language as the mode of communication.

For truly assistive, autonomous robots it will be necessary to implement the skills to adapt to new situations in complex dynamic outdoor environments and to communicate in a natural way with non-expert human users. Therefore theories from spatial cognition and computation need to be included in interactive robots.

1.3 Overview of Spatial Cognition and Computation

The subject of spatial cognition comprises cognitive functions that allow humans or animals to deal with spatial relations and orientations of objects in space, tackle visual spatial tasks, have an awareness of self-location, and solve navigational tasks. The equivalent of spatial cognition for robots and technical systems is spatial computation which can be implemented to model spatial cognition or provide technical systems with representations of spatial environments. These areas are important to this thesis as they provide methods and systems that inspire representation of and reasoning about spatial information.

1.3.1 Spatial Representations in Humans

One of the first researchers to study cognitive representations of large-scale spatial environments was Trowbridge. He used the term “imaginary map” [187] to describe an individual’s spatial representation of relations in the real environment. Tolman introduced the term “cognitive map” for the mental representation of the layout of one’s environment in humans and other animals [185]. Contrary to common belief at the time, he stated that complex behaviour, such as the learning of routes, cannot be explained by simple stimuli and reactions, but only by a mental representation of the routes under certain expectations. Later, additionally the term “mental map” was introduced by Gould and White [57]. The ability to construct and use different forms of spatial knowledge and acquire cognitive maps has been researched in children of different ages [139, 165] and in elderly people [42] as well. The cognitive map gives answers to where certain things are in one's environment and how to get to them from the current location. An operational definition of a cognitive map is given as "a representation of spatial relationships that enables computation of novel shortcuts between known locations" [132]. Characteristic attributes of cognitive maps are individuality, simplification of relations in the real environment, as well as a dependence on personal and social factors.

Cognitive maps are distorted representations of relations in the real world. Such distortions include the alignment and rotation of objects relative to each other [188] and geometrical simplification of objects [25, 120, 172] or distances [107]. Errors in cognitive maps can be caused by various sources, such as the general topography [27], barriers [85], the familiarity with an environment [18], or the number of turns or intersections on a route [155]. The representation of distances depends on proximity to subjective reference points [66]. Tversky [188] has discovered that distortions can be caused by the way humans encode spatial information into their memory which is based on cognitive hierarchical organization [117]. Individuals usually simplify the complex geometrical structure of their environment by representing it as a matrix of straight lines and right angle junctions [5], where the spatial extent of familiar and salient areas is exaggerated [58].

Cognitive mapping in spatial cognition was first described in [185] and is defined as the process by which spatial information about an environment is acquired, coded, stored, recalled, and decoded. The product of this process is a cognitive map. As humans have only limited sensing, information processing, and storage capabilities, cognitive mapping forms constricted and simplified representations of a complex, uncertain, and changing environment [40]. Cognitive mapping takes place in connection with the formation of travel plans which are adapted to facilitate travel [52]. It may even depend on an individual's personal value system [158].

Lynch [110] describes how humans perceive and organize spatial information in cities. He states that spatial knowledge consists not only of topological maps and geometrical models of the environment, but also of procedures for getting from one place to another. Lynch speculates that humans form mental maps of their environment, consisting of five elements, namely paths that can be navigated, edges representing perceived boundaries, districts as large areas with common properties, nodes such as focal points or intersections, and landmarks serving as reference points. By contrast, Golledge [56] supposes that an urban cognitive map is simply composed of a set of nodes and connecting paths. He points out that such a cognitive map develops over time as the individual gains experience.

Cognitive maps are the basis for wayfinding. The modern term wayfinding was coined by Lynch [110], who defined it as "a consistent use and organization of definite sensory cues from the external environment". Passini [136] expanded the term to include graphic and audible communication, tactile information, and logical spatial planning. Wayfinding can occur actively through navigating in an unknown environment, or passively by studying a map of an unknown environment before navigating.

The equivalent of spatial cognition in robots is spatial computation. A review of spatial computation is presented below.

1.3.2 Spatial Representations in Technical Systems

Spatial representations are the basis for navigation and path planning for technical systems. Autonomous robots require an internal representation of their spatial environment to be able to operate in it. Therefore "mapping is one of the core competencies of truly autonomous robots" [182]. Robot mapping is the action of acquiring a coherent representation of the surrounding environment. Spatial representations and mapping in robots are reviewed in [201]. A general problem of local robot mapping is that the robot needs to simultaneously localize itself within the environment while building a map of it. The general method that tackles this dual problem is Simultaneous Localization and Mapping (SLAM) [39, 121] which solves problems such as loop closure, uncertainty management, mapping and localization. The immediate environment can be mapped by robots based on scanning the surroundings with sensors, e.g. laser-range finders, sonar sensors, or camera systems. The raw sensor data is processed to extract relevant information and represent it as the most likely map of the real environment. The commonly used occupancy grid maps introduced by Moravec and Elfes [123] represent the environment by a grid of cells, each including the probability that the respective location is occupied by an obstacle. Grid based maps offer a relatively high resolution of the close environment and are the basis for local navigation and path planning. However, they are not suitable as long-term representations of dynamic large scale environments, where salient objects, relations and orientations between them, and passages are of interest.

Recently there are attempts to represent large scale spatial relations in such a way that not only geometrical information needed for navigation is stored, but also topological information for higher level operations such as task planning and HRI: Some approaches model route-based navigational knowledge of (indoor) environments such as route graphs [89, 199]. A robotic environmental model composed of different levels of abstractions, including a metric line map, a navigation graph, and a topological map is provided by a multi-layered conceptual map [210]. A similar approach introduces a multi-hierarchical semantic map for mobile robots [51] where spatial and semantic hierarchical information is linked. A spatial representation for a robot interacting with humans including metric, symbolic, and cognitive layers for individual robot and robot-human team behavior is presented in [78]. Hierarchical probabilistic representations of space based on objects are applied to equip a robot with spatial cognition [191]. Kuipers [91, 92] has developed a computational theory of the cognitive map, the Spatial Semantic Hierarchy (SSH). This theory was motivated by human spatial reasoning. It models the spatial belief of a technical system as a hierarchy of representations consisting of a control level, a procedural level, a topological level, and a geometrical level. The central element of the SSH is the topological layer which is logically prior to the metric representation. Recently, calculi for qualitative reasoning have been aplied to reasoning about spatial relations between objects and areas [32, 148]. Kuipers [93] also applied qualitative spatial reasoning to robot navigation and path planning.

Human-robot interaction as well as spatial cognition and computation are emerging research fields both in cognitive sciences and robotics. Combining knowledge from both of these areas can come by the challenges posed by enabling robots to extract representations of task information by asking humans for instructions.

1.4 Main Contributions and Outline of the Thesis

This thesis addresses methods for presenting robotic systems with abilities necessary for acquiring probabilistic representations of task information extracted from human-robot dialogs, based on the application example of extracting and representing missing route information. The thesis is organized into 6 chapters. The general outline of the main chapters of the thesis and the connections between the presented methods are visualized in Fig. 1.1. Chapter 2 presents an experiment with an interactive mobile robot. Research questions for the remainder of the thesis are deduced from the experimental results. In Chapter 3, guidelines for direction-inquiry dialogs between humans and robots are derived from linguistic principles and implemented in a dialog system. The dialog system constitutes the interface between human and robot and proactively extracts route information through HRI. Probabilistic models of route information, such as directions and distances are presented in Chapter 4. The probabilistic models serve as means to assess the reliability and accuracy of direction and distance information of individual route segments. Complementary to Chapter 4, Chapter 5 introduces a system for reasoning about overall route descriptions. If route information is assessed as plausible by this system, it is included in the route belief of a robot, i.e. the internal representation of the route information, otherwise the dialog system in Chapter 3 inquires about all conflicting information. The route belief can be augmented by metric sensor data during navigation. In the following the main scientific contributions of Chapters 2 to 5 are summarized.

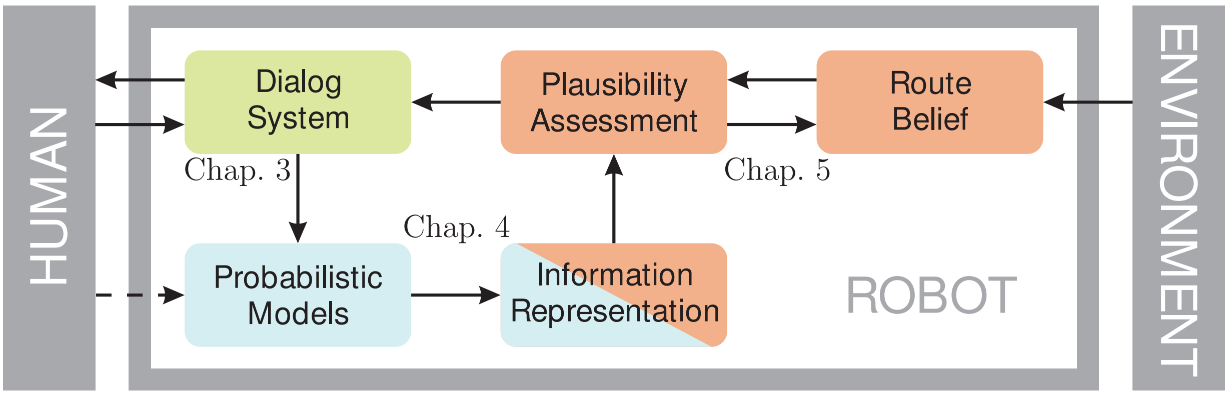


Fig. 1.1: Schematic outline of the thesis.

In order to define challenges and problems with which a robot asking humans for directions might be confronted, Chapter 2 describes an experiment that was conducted with the *Autonomous City Explorer* robot. The robot was developed at the Institute of Automatic Control Engineering of Technische Universität München for the task of autonomously navigating to a designated goal location without any previous map knowledge or GPS, but solely by asking passers-by for directions [224]. *Previous experiments by other researchers included either robots navigating autonomously through outdoor environments, or human-robot interaction in well-defined indoor environments. This experiment was the first to combine the challenges of autonomous outdoor navigation and HRI and enable a robot to navigate based on human route information.*

Research questions are derived from the results of the *ACE* experiment, namely proactive information extraction from human-robot dialogs, probabilistic representation of route information, and reasoning about different route descriptions. Methods and systems that answer these research questions are introduced in the following chapters.

Chapter 3 introduces a dialog system for robots inquiring for directions. In order to render the human-robot dialogs natural to humans and proactively extract unambiguous route information, the dialog system includes guidelines derived from human-human communication. For this purpose, principles from linguistics relevant to spatial discourse are reviewed and dialog guidelines are derived as policies for HRI. The dialog system is designed in a modular way to be extendable to other communication modalities and further discourse topics. An experimental evaluation shows that human participants subjectively rate the dialogs as natural and assess the applied guidelines derived from human-human communication favorably. At the same time the objective results of system performance are comparable to results from other researchers. *While other researchers focused on the technical aspects of natural language communication when developing dialog systems, this dialog system includes guidelines based on principles from human-human communication that render the dialog natural to humans and allow for proactively extracting unambiguous route information.*

Information in route descriptions given by humans can be distorted, simplified, and even erroneous. Therefore probabilistic models of direction and distance information for individual route segments in route descriptions are derived in Chapter 4. These models provide robots with means to assess the accuracy and reliability of such information. To assess the reliability of directions a probabilistic model for direction information assigns certainty values to individual route segments, based on the direction of the previous route and overall error probabilities. Posterior probability distributions of quantitative and qualitative distances provide means to assess whether locations in the real environments are referred to given corresponding descriptions. The effectiveness of the identified information models is visualized by representing an example route probabilistically. *Related work by other researchers provides only models for cognitive directions and distances which are different from the directions and distances given in route descriptions. This chapter does not only introduce probabilistic models for route description information that are novel to spatial cognition and computation, but has a meaningful application in robotics as well.*

Chapter 5 presents a system for *Simultaneous Reasoning and Mapping* (*SRAM*). The system includes the dialog system presented in Chapter 3 and the probabilistic models from Chapter 4 and presents a novel framework for representing, reasoning about, and storing route information. Extracted route descriptions are represented as route graphs and assessed for plausibility by comparing them with existing route belief using pattern matching and route similarity assessment. In this way rotations, gaps, or excess segments in route descriptions as well as distinct descriptions leading to the same goal can be identified. If necessary, the system causes the dialog system to inquire about conflicting information. A route graph that is assessed as plausible is combined with the existing route belief, while implausible information is discarded. Furthermore the route belief is augmented by sensor data during navigation. An experiment with human participants demonstrates the capability of the presented approach. *Reasoning about spatial concepts and relations* *has been studied by other researchers, however there is no state of the art work that tackles the problem of representing and comparing overall route descriptions. The presented SRAM system constitutes a novel framework for representing and reasoning about route descriptions coming from different sources, and building a route belief from them.*

Finally, Chapter 6 provides conclusions of the methods and approaches presented in the thesis and discusses possible directions for future research.

2 Identification of Research Questions from an Outdoor Experiment

An experiment where a robot was given the task to autonomously navigate to a designated goal location without any previous map knowledge or GPS, but solely by asking passers-by for directions is described. Research directions for the following chapters are deduced from the experimental results.

2.1 Problem Description and State of the Art

The goal of this thesis is to enable robots to construct route beliefs through human-robot dialogs as a basis for global robot navigation and path planning. An experiment was conducted in collaboration with several researchers[1] at the Institute of Automatic Control Engineering of Technische Universität München, in which a robot was given the task to reach the city center of Munich, starting at the university campus, without any previous map knowledge or GPS, but solely by asking passers-by for directions. On the basis of the experimental results, research questions that are answered in Chapters 3 to 5 are derived.

Robotics researchers have conducted experiments usually focusing either on autonomous outdoor navigation or on human-robot interaction. Among these are experiments on robust perception, navigation, and manipulation in everyday settings. Progress has been made in the field of unmanned outdoor navigation in unstructured terrains [184] and more recently in urban environments [21, 122]. However, all of these robots were provided global waypoints in the form of GPS coordinates, as well as topological information about the route in advance. These experiments did not involve any human-robot interaction. On the other hand, experiments have been conducted in the field of human-robot interaction in structured indoor environments. Prominent examples are robots as tour guides for museums [24, 181] and shopping malls [59] that successfully relay useful pre-compiled information to humans. These robots are typical service robots with classical roles, providing information to humans. However, they are not able to extract missing information and thereby adapt to unforeseen situations. A few experiments studied the problem of robots extracting information by asking humans for it. These include a space robot asking for information in cooperative manipulation tasks [46], a robot asking the way at a robotics conference [118], a miniature robot finding its way in a model town by asking for directions [101], and a robot that creates a map of its environment by exploring it and asking a human to label areas of interest [68]. However, all of these robots operate in structured indoor environments.

[1]The team in alphabetical order: Andrea Bauer, Martin Buss, Klaas Klasing, Kolja Kühnlenz, Georgios Lidoris, Quirin Mühlbauer, Florian Rohrmüller, Stefan Sosnowski, Dirk Wollherr, Tingting Xu.

In the presented outdoor experiment, the *Autonomous City Explorer* (*ACE*) robot is given no prior map knowledge or GPS, and has to ask passers-by for directions on the way in order to acquire the necessary direction information, build an internal representation of that information, and use it for navigation. The experimental results are positive in general, point out limitations of this system, and call attention to specific research questions that are answered in the remainder of the thesis.

This chapter is structured as follows. In Section 2.2 the *ACE* robot is introduced and a brief system description is given, while an extensive description can be found in Appendix A. Section 2.3 describes the experiment and discusses the experimental results and the limitations of the system. Open research questions are derived in Section 2.4. Finally, Section 2.5 provides a discussion of the experiment and the open research questions.

2.2 The *Autonomous City Explorer* Robot

The *ACE* robot comprises hardware and software for stereo image processing, interacting with non-expert human users, and autonomous outdoor navigation [214, 224]. *ACE* as depicted in Figure 2.1 is equipped with an active-stereo camera head for human tracking and gesture recognition, an animated mouth, a loudspeaker, and a touch screen for HRI, as well as a differential wheel mobile platform and laser range finders for navigation. The software of the robot is broken down into three main subsystems; the interaction system is presented in this chapter; while the navigation system is described extensively in [106]; and the vision system is presented in detail in [126]. More details on the hardware and software of *ACE* are provided in Appendix A.

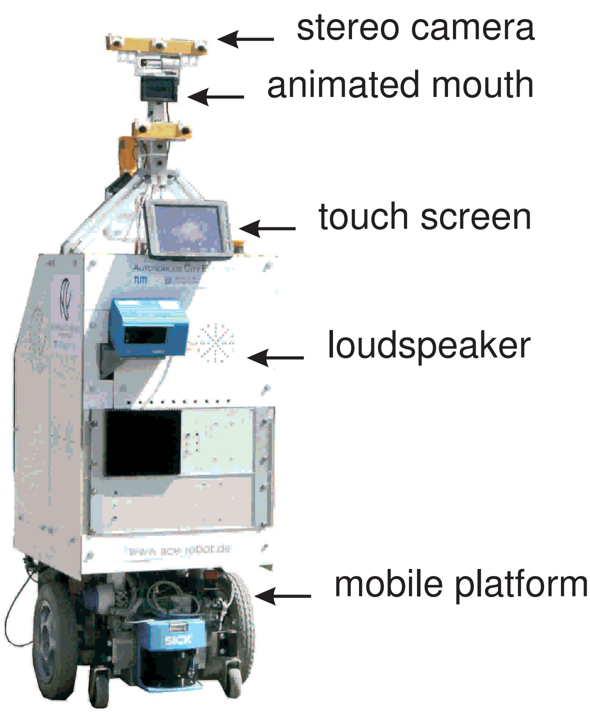


Fig. 2.1: The *ACE* robot with principal hardware components.

The robot communicates with humans by synthesized speech, using MaryTTS [162], which is augmented by a synchronously animated mouth displayed on a small monitor. Additionally the robot presents spoken text on a touch screen for robustness and convenience, along with informative images, e.g. camera views. As robustness to environmental disturbances such as noise is an important requirement for the system, the touch screen, is the main means of information input from the human. A graphical user interface (GUI), displayed on the touch screen, comprises buttons for possible answers and buttons that allow the user to change the language, go back to the previous step in the dialog, or quit the interaction. Speech is not included as a mode of input because speech recognition does not work robustly in outdoor environments, especially as traffic and human noises have the same frequency band as the interesting speech signals from the human partner. The robot recognizes and interprets pointing gestures, as described in [126]. These components constitute the basis for human-robot interaction in the experiment.

Human-Robot Interaction System

The interaction system is an integral part of *ACE*, as it communicates with humans, extracts route descriptions, and represents such information internally. As a starting point for the interaction system, a finite state machine (FSM) was developed on the basis of an interaction flowchart that includes all steps necessary for extracting route information. The FSM as the core component of the interaction system is responsible for interfacing with the hardware for communication and with the navigation and vision systems; it selects appropriate interaction behavior depending on the situation and controls human-robot communication. The FSM is depicted in Fig. 2.2, where the text the robot utters is noted under each respective state, marked by a speaker icon. Transitions between the states can be triggered by inputs from humans (marked by hand icons or boxed icons) or by signals from the navigation or vision system of the robot (denoted in grey).

During human-robot communication the robot asks the human to give directions to a designated goal location and extracts route information. Typical communications with *ACE* proceed as follows. Firstly, the robot addresses a human, introduces itself, and asks the human to give directions to the designated goal location. The robot asks the human to point in the direction it has to go first. The human is subsequently asked to give further directions through touch screen commands, where buttons for the four basic directions are provided. All direction information is depicted on the touch screen immediately. This allows the human to verify whether the robot has interpreted an input accurately and if necessary correct it. When the human has finished giving directions the robot thanks the human for the help and starts moving along the given route.

The information extracted from gestures and touch screen commands is represented internally as a route graph, similarly to [90, 199]. The route graph $G\langle N, E\rangle$ consists of nodes N_i representing intersections and edges E_k representing actions connecting the intersections. Edges store the topological spatial relations distance and direction relative to the last direction between intersections. The nodes in the route graph serve as global waypoints for path planning. During navigation the directions and distances of edges are updated by metrical data from on-board sensors.

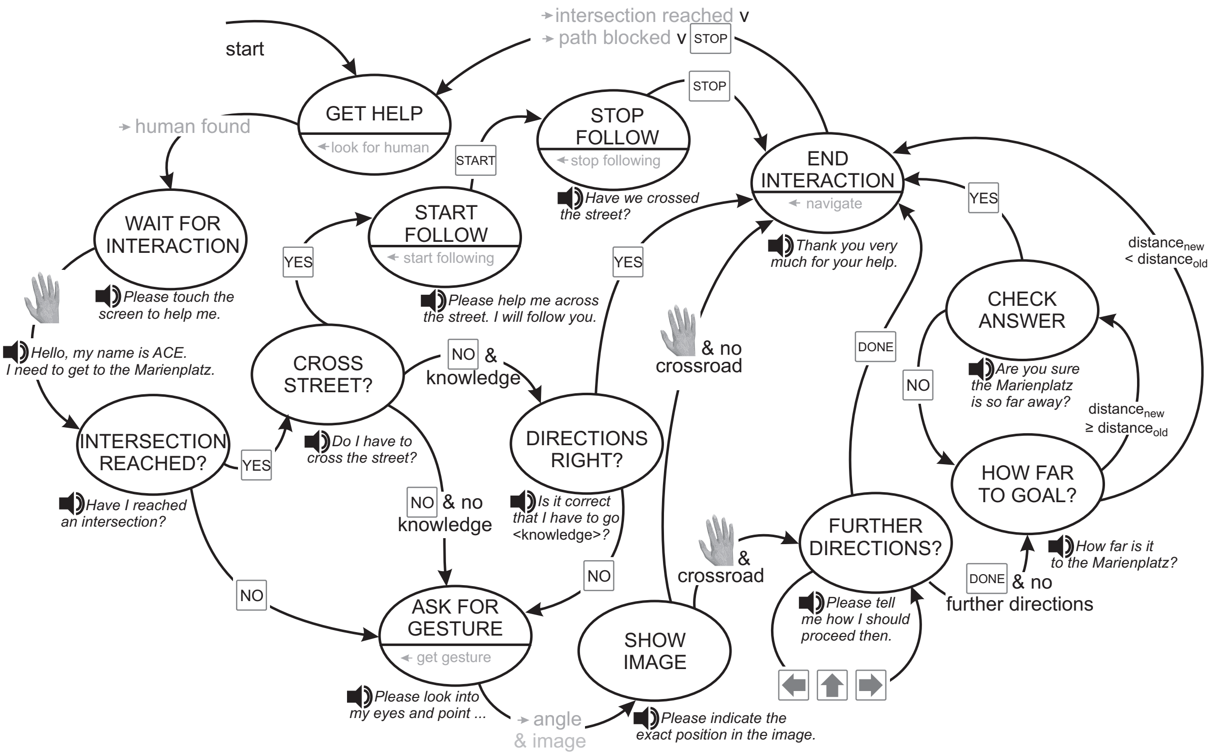


Fig. 2.2: FSM of the interaction module of the *ACE* robot. Transitions can be triggered by pressing buttons (boxed icons), by touching the screen (hand icons), or by signals from the navigation or vision system of the robot (grey text).

2.3 The *Autonomous City Explorer* Experiment

An outdoor experiment was conducted in the city of Munich, where the *ACE* robot had to reach Marienplatz, i.e. the central city square, starting from the campus of Technische Universität München. The robot did not have any prior map knowledge or GPS sensors, and therefore had to ask passers-by for directions in order to complete its task. The distance *ACE* had to cover was approximately 1.5 km, partly on sidewalks in a traffic zone and partly in a crowded pedestrian area.

The robot managed to travel the distance between the campus and Marienplatz in five hours. *ACE* interacted with 38 passers-by which explains in part the relatively long duration of the experiment. Many people stopped the robot on the way to interact with it out of curiosity. As *ACE* was designed as an interactive robot, the number of interactions was a positive sign, and not a limitation. In an application where the task of a robot is to reach a goal location efficiently, a robot would need to interact only as often as necessary.

The interaction partners were chosen according to random choice by the robot, and their willingness to interact with it. There were male and female interaction partners of all ages from children to elderly persons. The average duration of an interaction was 1.5 minutes. A snapshot of an interaction between *ACE* and a passer-by is shown in Figure 2.3. A passer-by points in the direction the robot has to go while *ACE* follows the gesture with its camera head to take an image and present it to the human.

Fig. 2.3: *ACE* following the pointing gesture of a passer-by with its camera head.

The fact that the robot reached its goal solely with the help of instructions from passers-by who were not previously instructed on how to interact, leads to the conclusion that the human-robot communication was successful. Difficulties arose where the human partners had too high expectations of the abilities of the robot. Many humans expected the robot to be able to understand speech at first and tried to answer through natural language until they realized that they had to use the touch screen to communicate. Also the robot had limited gesture recognition abilities when occlusions or inconspicuous pointing gestures occurred. This was compensated by the robot, by asking humans to specify their pointing gestures in a camera image presented on the touch screen. Limitations that can be identified here are the perception capabilities of *ACE* which arose from the requirement of robustness. In scenarios where robots rely on information given by humans, natural-language communication would be the interaction modality most natural to humans. Natural-language dialogs can also be designed to have a more flexible structure than that caused by the FSM in Fig. 2.2.

The social acceptance and people's willingness to support the *ACE* robot was investigated in collaboration with the ICT&S Center, University of Salzburg. The results of the survey conducted with bystanders and people who interacted with *ACE* [226] reveal that *ACE* achieved a high acceptability rate and that passers-by were willing to support the robot in its task. Furthermore, the interaction system was found to be intuitive above all to children, and supports short-term interaction in public space. Generally the participants stated that they did not have the feeling that they needed additional training to handle the robot, which shows that the chosen approach is intuitive to non-expert human users.

The route graph G built by the robot is depicted in Fig. 2.4 along with a corresponding occupancy grid in a satellite image of the real environment. The robot was currently positioned at node N_2, where it had extracted topological information of the route that lay ahead, shown as white lines. The part of the route the robot had already covered has been updated with metric data, shown as black lines. A limitation of this procedure is that the latest direction information is presumed to be the most reliable and therefore is followed by the robot during navigation. The system does not compare the route information to information by other humans in order to assess it for plausibility.

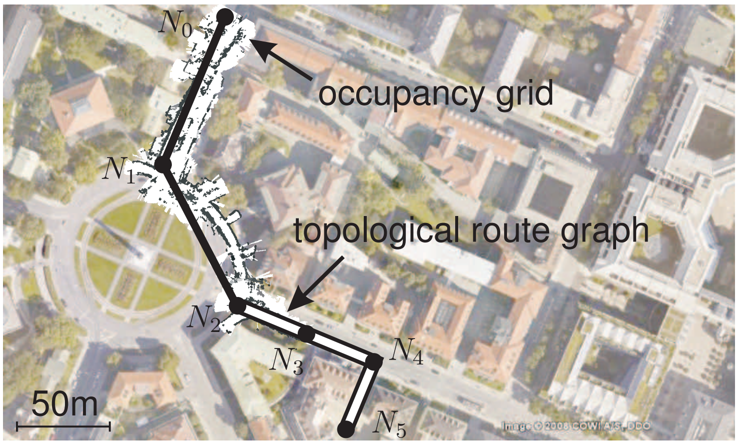


Fig. 2.4: Example of a route graph extracted by *ACE* from human instructions.

The robot was sent in the wrong direction by a passer-by once. This wrong information was corrected by the next interaction partner when the robot stopped again. A clear limitation of the interaction approach in the *ACE* robot is that direction information is not assessed for plausibility in order to identify erroneous information.

The experimental results are positive in general with a few limitations. From these, open research questions can be derived that are answered in the remainder of the thesis.

2.4 Open Research Questions

The results of the *ACE* experiment confirm the appropriateness of the approaches for extracting route information for navigation from human-robot dialogs, while at the same time they point out some limitations of the implemented approaches for HRI. Based on these, research questions are identified that are answered in the following chapters.

Research questions are derived from the limitations in the results of human-robot communication, information verification, and route belief building. The research questions that need to be answered in order to enable a robot to extract route information from HRI and represent it for navigation are:

- How can robots proactively extract route information from natural-language dialogs?
- How can robots identify and represent inaccuracy in route-description information?
- How can robots build a plausible route belief from extracted information?

These research questions are answered in the remainder of the thesis. Chapter 3 presents a natural-language dialog system for robots extracting route information proactively from human-robot dialogs. Probabilistic models for individual direction and distance information are derived from route description data sets in Chapter 4. Finally, Chapter 5 presents a system for simultaneously assessing route descriptions for plausibility by comparing them to other descriptions and building a route belief from plausible information.

2.5 Discussion

In order to be truly assistive in intelligent and versatile ways robots need the abilities to extract missing task information, such as route descriptions, from HRI and represent it in an internal representation. In order to define relevant research questions about these abilities, results from an experiment are analyzed where a robot is given the task to navigate to a designated goal location in an unknown outdoor environment by asking passers-by for directions. The *Autonomous City Explorer* robot is presented with a focus on the implemented interaction system. A description of the experiment with a discussion of the experimental results is given and open research questions are identified.

Experiments by other researchers included either robots navigating autonomously through outdoor environments, or human-robot interaction in well-defined indoor environments. The experiment described above presented a novel robotic system capable of reaching a specified goal location in an unknown, complex, and dynamic urban environment by extracting route information from passers-by.

The success of this experiment suggests that the steps taken towards proactively extracting route information from HRI were appropriate in general, while there was some room for improvement. Research questions are deduced from the experimental results and the limitations of the system. They are answered in the following chapters.

Many humans expected the robot to be able to communicate through natural language and tried to talk to the robot. This shows the need for a natural-language dialog system that extracts route information proactively. As natural language is often vague and ambiguous, such a dialog system needs to adapt mechanisms from human-human communication to compensate these challenges. Chapter 3 reviews principles from linguistics relevant to spatial discourse, derives guidelines for HRI, and presents a dialog system for proactively extracting route information from natural-language dialogs.

The fact that the robot did not verify extracted route information and consequently was sent in the wrong direction once, indicates a need for methods to reason about given information and to assess route descriptions for plausibility. Both, individual sections of route descriptions and whole route descriptions can be inaccurate or even erroneous. Therefore, methods for representing route information probabilistically and reasoning about different route descriptions need to be developed. Chapter 4 solves the problem of possible errors or inaccuracies of individual route segments by presenting probabilistic models for direction and distance information in route descriptions. These probabilistic models can be used by robots to assess the reliability and accuracy of individual route information. Chapter 5 presents a system for reasoning about route descriptions. In this system extracted route descriptions given by different humans are compared and assessed for plausibility while simultaneously building a route belief as an internal representation of the plausible route information. This system interfaces with the dialog system in Chapter 3 and applies the probabilistic models in Chapter 4 to individual route information.

3 Human-Robot Dialog for Route Information Extraction

This chapter presents aspects of human-robot communication necessary for extracting a route belief. Principles from linguistics relevant to spatial discourse are discussed and guidelines for human-robot direction-inquiry dialogs are derived. Based on these guidelines, a dialog system is developed that allows for natural language communication between humans and robots as well as for proactive route information extraction.

3.1 Problem Description and State of the Art

The first step towards building an internal representation of information about an unknown environment is to access this information. A legitimate approach for robots to come by missing knowledge about complex and dynamic environments is to ask humans for information. As natural-language is the most common and intuitive mode of communication for humans, natural-language communication is a necessary requirement for this task and has to be included in robot dialog systems. Considering that "vagueness is one of the most salient, but also one of the most effective features of natural language" [82], dialog systems for robots have to cope with possible ambiguities and interpret information extracted from dialogs with humans correctly. Therefore it is essential to include findings from linguistics relevant to the respective discourse topic in robot dialog systems.

As a specific example of acquiring missing task information through HRI, this chapter concentrates on robots asking humans for directions. The ability to ask for directions gives robots the competence to react quickly and flexibly to changes in task requirements or the environment. Furthermore, spatial discourse in general and route descriptions in particular provide a well-defined contextual field with all the challenges of natural language processing, such as ambiguity, vagueness, and contextual dependency.

Dialog systems are being developed for various technical applications and contexts. Extensive surveys of spoken language dialog systems are presented in [72, 168]. As one of the first dialog systems ELIZA was developed by Weizenbaum [198]. Similar techniques are still used in chat-bots, as e.g. presented in [41, 141], that try to carry on a conversation without understanding the given information. Currently deployed dialog systems include information for travelers calling airlines, car rentals, and other travel providers; automatic speech recognition and text-to-speech systems in cars; and interactive virtual agents serving as tutors for children [34]. Exemplary dialog systems with spatial content are a multimodal urban information system [60], and dialog systems for information about public transport [147, 174].

Robotics researchers have understood the need of robots to exchange spatial information with humans. A robot that extracts spatial information from a grid map and translates it into linguistic spatial descriptions is presented in [167]. Robots that create maps from sensory information and asking humans to label places and areas of interest are presented in [68, 171]. These robots only use HRI as a means to assign human understandable labels to regions and features in their spatial belief. Deictic communication between humans and robots has been investigated in [175], where a human pointed out certain objects in a room to a robot by speech and gesture. An approach for conversational mobile robots that can give spatial information to humans is presented in [209]. MACK, an embodied conversational kiosk that can provide humans with information and give directions via speech and gestures is described in [28].

Robots that ask humans for directions in order to extract information about their environments mostly still operate in very simple structured indoor environments. Coarse qualitative route descriptions can be understood by some robots [8, 118, 119]. However, these robots cannot cope with the complexity and vagueness of natural language. Wei et. al. have introduced a global inference approach [197] that aims at having a robot automatically find a path within an office environment based upon human directions. A miniature robot that can find its way in a model town by asking for directions is described in [101]. An action inference approach [112] was found to improve task success of a virtual agent following human route descriptions, compared to following purely explicit instructions which indicates that a deeper understanding of implicit information in route directions is indispensable. A situation with reversed roles is presented in [133], where a robot gives route descriptions via speech and gestures in a shopping center. All of these studies concentrate on robots exchanging route descriptions or spatial information through HRI in structured and mostly static indoor environments.

As the ability to ask for directions is expedient for robotic systems, while route descriptions may be vague and ambiguous, the need arises for linguistic guidelines for direction-inquiry dialog systems. These will facilitate human-robot dialogs and enable robots to reason with humans about the extracted spatial information and interpret the route information correctly. For that purpose linguistic principles relevant to human-human spatial discourse are reviewed, and guidelines for human-robot dialogs are deduced. A dialog system implementing these guidelines is described here. This dialog system constitutes a basis for extracting probabilistic representations of route information from human-robot dialogs. Probabilistic models of extracted individual route information in route descriptions are presented in Chapter 4. A system for comparing and reasoning about different route descriptions is introduced in Chapter 5.

The remainder of this chapter is structured as follows. In Section 3.2 linguistic principles relevant to direction-inquiry dialogs are reviewed. Based on these guidelines for human-robot dialogs are derived in Section 3.3. A dialog system implementing these guidelines is presented in Section 3.4 and evaluated experimentally in Section 3.5. Finally, Section 3.6 provides a discussion of the experimental results and the contribution of this chapter.

3.2 Linguistic Principles Relevant to Direction Inquiry

Linguistic theories in general deal with the structure of complex verbal actions within human-human communication and can therefore serve as models for HRI. Two of the most prominent and relevant works in linguistics are those of Chomsky [30] who introduced formal description tools for various languages and Fillmore [43] who introduced semantic cases as a form of semantic grammar opposed to syntactic grammar. These works represent the groundwork for computational linguistics as well. Other fundamental concepts in human-human communication include turn-taking [154] which consists of a turn constructional and a turn allocational component, and grounding [31] which is the collective process by which the communication partners try to reach a mutual belief.

This section reviews linguistic principles pertinent to spatial discourse. Principles for reasoning about space can be found in linguistics including the analysis of direction-inquiry dialogs and of the different semantic meanings of verbal expressions or gestures. Theories relevant to the problem of asking for directions are the analysis of dialog structures and the complex deixis theory founded by Bühler [23].

3.2.1 Dialog Structure

Wunderlich [203] analyzed direction-inquiry dialogs and identified a common structure of four consecutive phases:

I *Introduction*: The questioner addresses a respondent and defines the task, i.e. giving directions to a specified goal location, possibly defining the mode of transportation or other individual requirements.

II *Giving directions*: The respondent provides the necessary information by means of natural language and gestures, sometimes additionally with the help of a sketch.

III *Confirmation*: Either of the two partners confirms the information. Further inquiries and corrections can be made.

IV *Conclusion*: The questioner thanks the respondent and they part.

This schematic structure is very flexible, i.e. some phases may be interchanged or recur. Nevertheless it is a well-proven guideline for human-human dialogs reflecting the intrinsic cognitive processes involved. One of these cognitive processes for the respondent is planning the description by building a cognitive map [110, 165, 185, 187] which is based on individual experiences. This cognitive map includes "objects which are salient landmarks for nearly everybody" [81]. The respondent has to complete the task of separating these salient landmarks from individual experiences and present them to the questioner. If this is achieved by using appropriate means of communication, the questioner is able to build a corresponding cognitive map that represents the route to the goal, structured by landmarks.

Spatial information in general and route descriptions in specific are communicated using deictics, or deictic words, which are analyzed by deixis theory.

3.2.2 Deixis Theory

Deixis theory [23] is based on the assumption that communication acts can be assigned to two different fields of language, namely the *symbolic field*, comprising nouns which are symbols independent of the context; and the *deictic field*, including relating expressions, i.e. deictics, which vary depending on the context. In natural language communication referring is generally managed by deictics which point to subspaces within the deictic field and are the verbal equivalents to pointing gestures. They can be accompanied by gestures and/or movement verbs. In order to point to those subspaces both speaker and listener have to share a common "deictic space". At the beginning of a face-to-face route description this deictic space is given through the range of the visual perception. In the course of the conversation the route description usually leaves the shared visual perception range and a new deictic space is built by the geographical knowledge of both partners. Another deictic space can be introduced by involving a map or a sketch that represents the real geographic space.

Referring to subspaces in a certain deictic space by using deictics depends on particular contextual factors such as the position of the speaker and the direction of gaze [81]. These factors form a personal reference system.

3.2.3 The Origo

Bühler introduces the term "origo" [23] which is conceptually conceived as the point of origin of a "coordinate system of subjective orientation". It is derived from the need of a "basic reference point" [81] in a given deictic space which includes three deictic dimensions [3]: the *personal dimension* including personal pronouns; the *spatial dimension* including spatial demonstratives, adverbs, prepositions, and movement verbs; and the *temporal dimension* encompassing temporal expressions. Thus the origo, as the point of origin of the personal reference system, is defined by the personal mark *'I'*, the spatial mark *'here'* and the temporal mark *'now'*. Accordingly the origo is the initial reference point for personal, spatial and temporal deixis which relates to other elements within these three deictic dimensions. Thus deictics can only be interpreted relative to the origo. Deictics have certain roles within a direction-inquiry dialog depending on their deictic dimension:

- *Personal deictics* define the actors within communication.
- *Spatial deictics* indicate the directions from one decision point to the next.
- *Temporal deictics* refer to the time domains of actions which have to be carried out to reach one decision point from another.

The roles of deictics can be refined further on the basis of contextual factors and relations, as described by Fillmore [44] and Lyons [111].

In some cases deictics can only be located by using an additional "relatum", an object which is related to the origo [50]. For instance if a subspace is referred to relative to a landmark, the landmark functions as a relatum which is given relative to the perspective of the origo. However, if it is referred to a subspace relative to the speaker's body, then the origo, or the speaker, functions as a relatum. If not given explicitly in a route description the last decision point is the relatum.

Problems and difficulties in interpreting deictics can occur, as deictics can be interpreted differently according to the deictic dimension, the context or whether or not they are used in combination with a relatum.

3.2.4 Problems in Interpreting Deictics

Several problems and difficulties have been identified by Klein [81], and may occur while interpreting deictics and identifying the subspaces to which they refer.

- *Coordination problem:* In a dialog situation each participant has her own origo. Since the origo is usually defined by the position and orientation of the current speaker, the listener must project that origo into her own system of orientation. As soon as the roles of speaker and listener are exchanged, the origo of the new speaker becomes essential and the other person has to adapt to it. Sometimes it is not clear whether the origo in a route description is the one of the speaker or the one of the listener. In the case of a direction-inquiry dialog the origines of the speaker and the listener are the same in terms of the position, as *'here'* encloses both speaker and listener. However, the orientations of the two communication partners differ. This leads to ambiguous deictic meaning.

- *Problem of the shifted origo:* In the course of the direction-inquiry dialog both speaker and listener shift their origines into the perspective of an "imaginary walker" [81] representing the addressee on her way projected into the future. A place that would be normally referred to as *'there'* may be called *'here'* within the route description, e.g. *'go straight until you see the park, here you need to turn left'*. Thus the deictic *'here'* does not necessarily refer to the actual location but to the position of the shifted origo depending on the context.

- *Problem with the use of an analogon:* When humans use a sketch or a map to illustrate the described route, they introduce a new deictic space. Consequently there are two deictic spaces involved, the map and the real geographic space represented by the map. The map functions as an *analogon*, where pointing to an element within it represents pointing to an element in the real space. Problems arise when the assignment of the two deictic spaces is not clear.

- *Delimitation problem:* The subspace of the deictic space that a deictic points to cannot be fully identified simply by coordinating the origines of the dialog partners. The extent of a subspace is often vague and depends on context and environment, where the borders of the subspace must be established by gestures, verbal explanations or factual knowledge. For example, *'here'* can only be characterized as a subspace of the deictic space including the origo.

Even for humans it is sometimes hard to interpret deictics accurately, as any of these problems may occur in dialogs. For robots it is even harder, as they have to infer the intended meaning from a series of keywords extracted by their speech recognition systems.

In the following guidelines are proposed in the form of guidelines that can be implemented in robotic systems to circumvent or solve the above problems in HRI.

3.3 Guidelines for Human-Robot Communication

From the linguistic principles reviewed above, guidelines for HRI are deduced that can be seen as policies for dialog systems for robots asking for directions. These guidelines are summarized below and should in hypothesis make a dialog more successful in terms of rendering it more natural and intuitive for the human partner, enabling the robot to interpret the possibly vague deictics unambiguously, and extract unambiguous route information. The identified dialog guidelines (DG) are adopted from mechanisms in human-human communication reviewed above and specify requirements for technical systems.

DG 1 *Transfer of dialog structure from human-human communication*

This guideline should in hypothesis render the interaction natural and intuitive for humans, as the dialog structure is familiar to them from human-human communication. Therefore the dialog is structured according to the four phases *Introduction*, *Giving Directions*, *Confirmation*, and *Conclusion*, as shown in Fig. 3.1.

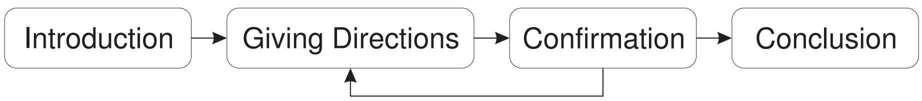


Fig. 3.1: Typical structure of direction-inquiry dialogs.

DG 2 *Initial alignment of the personal reference systems*

When directions are given it is not immediately clear whether they are given relative to the personal reference system of the human or the robot. As outlined in Fig. 3.2, a point of interest is described as being located in opposite directions relative to the reference system of the human O_H and the robot O_R, rendering the information ambiguous. The human can be asked for example to point in the first direction of the route to solve this problem, as a gesture is unambiguous. In mathematical terms, the pointing introduces a common reference system O_{HR} which is oriented towards the indicated direction, and specifies the necessary transformations, ${}^{R}T_{HR}$ and ${}^{H}T_{HR}$ of the reference systems O_H and O_R. This guideline solves the coordination problem and is crucial for the success of the whole dialog.

DG 3 *Internal representation of the route as route graph*

The robot needs an internal representation of the route information to be able to reason about it and to navigate by it. Therefore the perceived route description is broken down into route segments and stored as a topological route graph.

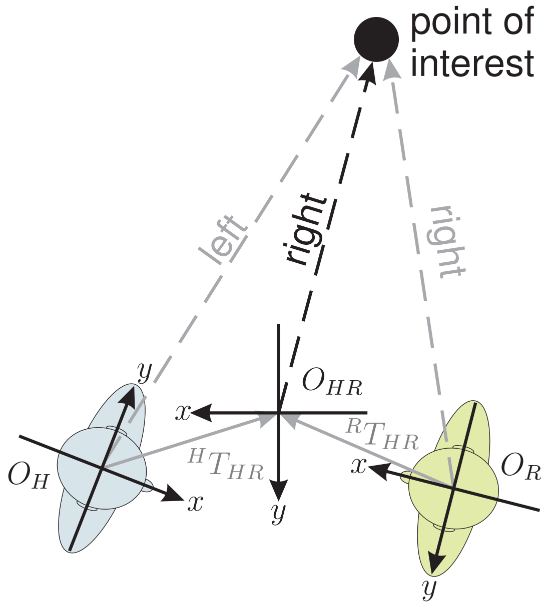


Fig. 3.2: A point of interest is referred to differently from different personal reference systems O_H and O_R. After coordinate transformations, $^{R}T_{HR}$ and $^{H}T_{HR}$ respectively, both dialog partners use a common reference system O_{HR} and refer to a point of interest in the same way.

DG 4 *Interpretation of basic directions*

The deictics *'left'*, *'right'*, *'straight'*, and *'back'* are the most important and basic information in a direction-inquiry dialog; interpreting these is the minimal requirement for a dialog system. These deictics represent actions and constitute the edges in a route graph.

DG 5 *Identification of explicitly given landmarks*

Landmarks, or decision points, structure the route description by providing start and end points of route segments. They correspond to nodes in the route graph, and must be identified when they occur in the route description.

DG 6 *Interpretation of implicitly given landmarks*

To solve the problem of the shifted origo, it has to be considered that the deictics *'here'* and *'there'* are usually linked with explicitly given landmarks, i.e. they either stand in the place of landmarks or accompany them. Thus they are implicit landmarks, and structure the route description just as other landmarks do. They are represented internally as nodes in the route graph.

DG 7 *Identification of distance information*

Distance information in a route description provides further information on actions that connect nodes in the route graph. This information must be added to the route belief and ideally be evaluated for accuracy by a probabilistic model (see Chapter 4), as distance information is only an estimation by the human and not an accurate measurement.

DG 8 *Differentiation between different types of movement verbs*

Movement verbs can be differentiated as translational or rotational verbs. Translational movement verbs define actions that connect the landmarks and must be identified by the robot. On the other hand verbs describing rotational movements denote a change in the orientation of the origo and are necessary for the interpretation of subsequent information.

DG 9 *Storage of deictics with relevant semantic attributes*

In order to simplify and disambiguate the interpretation of deictics, their relevant semantic attributes must be stored in the dictionary of the dialog system. This guideline helps solving the problem of the shifted origo, as deictics can only be interpreted unambiguously in relation to the context. Semantic attributes of deictics are part of the context. They can be the *deictic dimension*, the *distance range*, the *delimitation*, and whether or not a *relatum* is needed. Table 3.1 gives an overview of the most common deictics with the relevant semantic attributes; it does not provide an exhaustive list of deictics, but gives the reader an idea of how to classify deictics.

Tab. 3.1: Common deictics with relevant semantic attributes.

deictic	dimension			distance to origo			delimitation		relatum	
	spat.	temp.	pers.	inclusive	close	far	yes	no	yes	no
'here'	×			×				×		×
'there'	×				×	×		×		×
'close'	×				×			×	×	
'far from'	×					×		×	×	
'left'	×							×	×	
'right'	×							×	×	
'in front of'	×							×	×	
'behind'	×							×	×	
'this'	×				×		×			×
'that'	×					×	×			×
'now'		×		×				×		×
'then'		×			×	×		×	×	
'soon'		×			×			×		×
'later'		×				×		×		×
'I'			×	×			×			×
'you'			×		×	×	×	×		×

DG 10 *Mapping of temporal to spatial domain*

In a route description the passing of time corresponds to traveling along the route. Thus time and space have a similar structure and can be mapped to one another. In this way the number of expressions to be analyzed is reduced and the interpretation of the route description is simplified.

DG 11 *Modeling of distance ranges of spatial deictics*

Spatial deictics such as *'here'*, *'close'*, and *'far'* define different distance ranges. It is only clear that the deictic *'here'* denotes the position of the actual or shifted origo and that *'close'* and *'far'* do not include the origo. However, the actual extents of the regions referred to depend on several contextual factors. In order to solve the delimitation problem the corresponding distance ranges must be estimated using models found through user studies, as presented in Chapter 4.

DG 12 *Clear assignment of sketch and real environment*

A sketch depicted by the robot while a route description is given can be assigned unambiguously to the real environment by depicting the route with directions and relations between landmarks. The problem with the use of an analogon is solved in this way. An assignment to the real environment is given through the shared perspective of the common reference system both dialog partners use.

These dialog guidelines solve the problems that can occur in interpreting deictics and present policies for robot dialog systems for asking humans for directions. These guidelines are implemented in a dialog system as presented below.

3.4 Dialog System for Proactive Route Information Extraction

A dialog system is developed that proactively extracts route information from direction-inquiry dialogs. It is based upon the proposed guidelines derived from linguistics theories and designed in a modular way to be able to cope with multiple communication modalities and be extendable to various discourse topics. The dialog system guides the human through a direction-inquiry dialog and extracts all necessary route information. The implemented guidelines ensure that the dialog is natural to the human partners and that unambiguous information is extracted. In the following the structure of the dialog system and the integration of the proposed guidelines are presented.

3.4.1 System Overview

The architecture of the dialog system depicted in Fig. 3.3 consists of four main components: the User Interface, the Dialog Manager, the DataBase, and the Route Representation. The User Interface provides the communication port between the human user and the system. The Dialog Manager is the core component of the dialog system. It processes the input provided by the User Interface, extracts the relevant information, and generates feedback and answers. The proposed guidelines are implemented in the Dialog Manager and as dialog policies in the Database. The data necessary for processing the information is stored in the Database. It includes lists of associated keywords, general keywords, and dialog policies. The route information extracted from the dialog is stored as a route graph in the Route Representation of the robot.

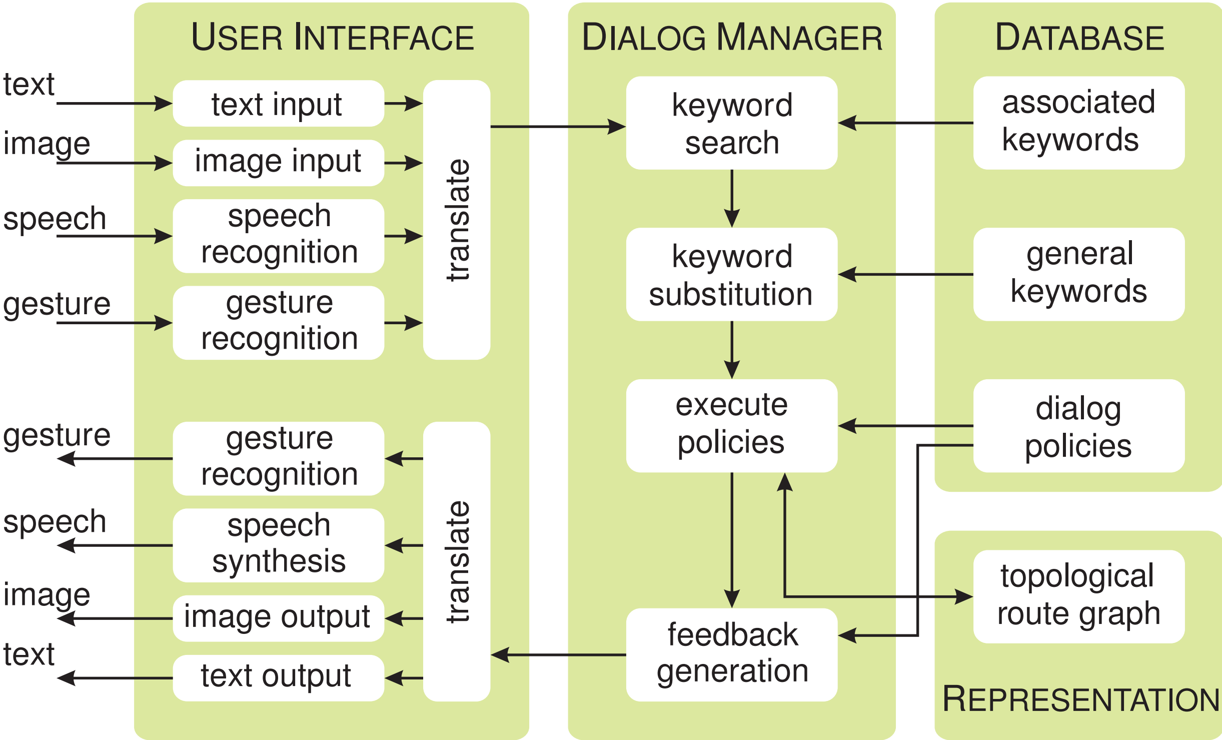


Fig. 3.3: The architecture of the dialog system.

Connection to Humans: The User Interface

The User Interface represents the connection between the human and the dialog system. It receives and pre-processes information from the human partner. Possible modes of communication are text, images, speech, and gestures. Perceived information is translated into text or symbols and a tag for the modality is attached. The input text is forwarded to the Dialog Manager. For example, a pointing gesture is translated into a direction, while a simultaneously given deictic is translated into text. Both information units are processed jointly in the dialog manager.

The feedback generated in the Dialog Manager is translated into corresponding modalities and subsequently presented to the user by the User Interface.

Core of the System: The Dialog Manager

The Dialog Manager is the core component of the dialog system. Based on DG 1 a finite state machine structures the dialog into dialog phases as depicted in Fig. 3.4. The FSM includes the four dialog states Intro, GivDir, Conf, and Concl representing the dialog phases *Introduction*, *Giving Directions*, *Confirmation*, and *Conclusion*, respectively. The additional system state Idle is active when no dialog is in process.

The system is initiated in the Idle state. If the system has no or only an incomplete route belief $\overline{B}_R$, the FSM switches into state Intro and begins a dialog with a human by introducing itself and the task. If the human has a belief B_H about the route and is willing to share it, the transition to state GivDir is made. Otherwise $(\overline{B}_H)$ the system switches to state Concl and ends the dialog. When the human is done giving directions,

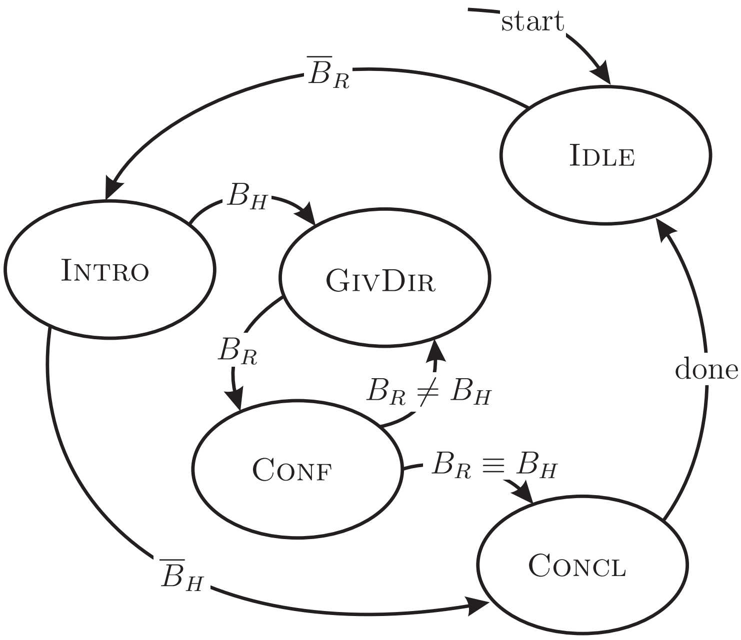


Fig. 3.4: The FSM of the Dialog Manager with five states controlling the dialog phases.

the robot saves the route information as belief B_R and the transition to state Conf is made where the extracted route information is confirmed. Should the belief of system and human differ, the system switches back to the GivDir state for grounding. During the Conf state the system recapitulates the extracted route information segment wise and asks the human to confirm each segment. If the belief B_R and B_H correspond ($B_R \equiv B_H$), the system switches to state Concl. In the case that the human disagrees with a route information ($B_R \neq B_H$), the FSM switches back to the GivDir state to allow the human to correct the respective route segment by providing new route information. Similarly, if the direction in a route segment is not included in belief B_R, a transition to GivDir is made and the human is asked for the missing direction. After each correction the system changes to Conf again, and finally to state Concl. After the dialog is concluded in state Concl the transition to state Idle is made.

The input from the human is processed by the Dialog Manager as follows. First the input string of words is compared to a list of keywords, i.e. associated keywords, in the Database. The associated keywords identified in this way are replaced by corresponding high-level keywords, i.e. general keywords. Thus a list of general keywords is assembled from the input information. In the next step the general keywords trigger the appropriate dialog policies for each state, stored in the Database. On the one hand dialog policies effect that the route information extracted in the GivDir state is forwarded to the Route Representation to be stored. On the other hand they permit the generation of appropriate feedback which is sent to the User Interface to be presented to the human.

Dialog Content: The Database

The Database contains associated keywords, general keywords, and the dialog policies as Extensible Markup Language (XML) files which can be extended easily. The associated keywords include all expressions that can be recognized by the system. There can be several associated keywords with the same meaning. General keywords abstract associated keywords with the same semantic content on a higher level, i.e. all synonymous associated keywords map to one general keyword. The Database contains dialog policies which define actions triggered by general keywords or by other policies. These actions include storage of extracted information in the route belief and feedback to the human partner. Most of the presented guidelines are implemented as dialog policies, as described below.

Representation of Information: The Route Belief

As postulated by DG 3 the extracted route information is represented internally as a topological route graph in the Route Representation of the system. Information in the topological route graph consists in directions, landmarks, and distances extracted from the dialog, as proposed by DG 4, DG 5, DG 6, and DG 7.

The topological route graph $G\langle N, E\rangle$ includes nodes $N_i(l_i)$ representing landmarks along the route of type l_i, and edges $E_k(N_i, N_j, \delta_k, d_k)$ representing actions connecting the landmarks. Edges hold information about the direction δ_k from node N_i to N_j, i.e. the input direction information is represented as an angle relative to the previous direction, and the distance d_k between the nodes N_i and N_j. The landmark types l_i and the distances between the landmarks d_k are optional information, i.e. they are not necessarily provided by the human, and if missing are represented by default values. A more extensive description the construction of topological route graphs and an approach for reasoning about them is given in Chapter 5.

3.4.2 Implementation of the Dialog Guidelines

The guidelines for HRI derived from linguistic principles relevant to direction inquiry have been implemented in the dialog system.

An FSM structures the dialog according to DG 1, as depicted in Fig. 3.4. The FSM includes the four states Intro, GivDir, Conf, and Concl. Before asking for further route information the human is asked to point in the first direction according to DG 2. This has been implemented in the *ACE* robot as described in Chapter 2, where the robot found the way to a designated goal location without previous map knowledge or GPS, but by asking humans for directions. The direction extracted through the gesture recognition system of the robot is forwarded via the Dialog Manager to the Route Representation to be stored as the first direction information.

A topological route graph is constructed according to DG 3 in the Route Representation component of the system. According to DG 4 the robot recognizes directions in the input text, and includes them in the topological route graph as direction δ_k of edge E_k. Landmarks are recognized and forwarded to the Route Representation component as suggested by DG 5 and integrated in the topological route graph as node $N_i(l_i)$. When a landmark type is not given explicitly, a default type, i.e. intersection, is inserted for l_i according to DG 6, when associated keywords representing landmarks are found, e.g. *'there'*, *'then'*. A representation of the extracted route can be depicted by the system including spatial relations between landmarks, such as directions and distances, according to DG 12.

Movement verbs are identified according to DG 8. Most movement verbs can be mapped to the general expression *'move'* which defines the action along an edge in the route graph. If no verb is given explicitly, *'move'* is inserted. The verb *'turn'* is not mapped in this way, as it does not define an action along an edge, but a change of the orientation, i.e. the origo, at a certain node, therefore defining a direction.

As suggested by DG 9, relevant semantic attributes of deictics are stored in the database within the dialog policies. This for example includes analyzing the meanings of deictics that need a relatum together with the reference object, or interpreting spatial relations according to the distance to the origo and the delimitation of deictics.

According to DG 10, temporal deictics are mapped to spatial expressions. Deictics accompanying or representing landmarks, e.g. *'then'*, are treated as their spatial equivalents, e.g. *'there'*. Accordingly, these temporal deictics are substituted by the same general keywords as the respective spatial deictics. Information about remaining walking time is converted to the equivalent walking distance, assuming a constant mean walking velocity.

The spatial extents of the non-delimited spatial deictics *'here'*, *'close'*, and *'far'* are derived probabilistically from a model presented in Chapter 4 according to DG 11, to facilitate the interpretation of those deictic words and obtain an assessment of distance d_k.

DG 7, stating that distance information should be recognized by the system, was implemented in such a way that distance information is allocated to the respective edge E_k and represented as distance d_k. If no distance is provided explicitly, a default value is inserted.

The next section presents an experimental evaluation of the dialog system including the dialog guidelines derived from linguistics.

3.5 Evaluation

An experimental evaluation of the dialog system has been conducted in order to assess the objective measure of system performance and the subjective measure of user satisfaction. The dialog guidelines are evaluated subjectively by the human subjects as well.

3.5.1 Experimental Setting

The dialog system has been evaluated by 20 undergraduate university students, 7 female and 13 male, aged between 20 years and 49 years, with a mean age of 27 years. The subjects interacted with the dialog system running on a PC in a laboratory to ensure equal experimental conditions. The input modality was restricted to text to ensure comparability between subjects and to interaction in outdoor environments, as speech recognition is to date still highly dependent on the speaker and on environmental noise. The graphical user interface used in the experiment was text-based where the subjects could type in their utterances. To simulate imperfect recognition, as it occurs in all speech recognition systems and even in human speech perception, the text was not pre-processed in terms of a spell check, but processed raw including typos or colloquial expressions that were unknown to the system. Additionally the GUI showed a simulated urban environment to give the subjects an overview of the simulated situation. The GUI is depicted in Fig. 3.5. The robot position and orientation within a map was predefined and had to be employed as the initial reference system when describing the route. The GUI included an interactive part, with text input, text output, and a display area for a schematic map of the elicited route believe B_R which could be hidden.

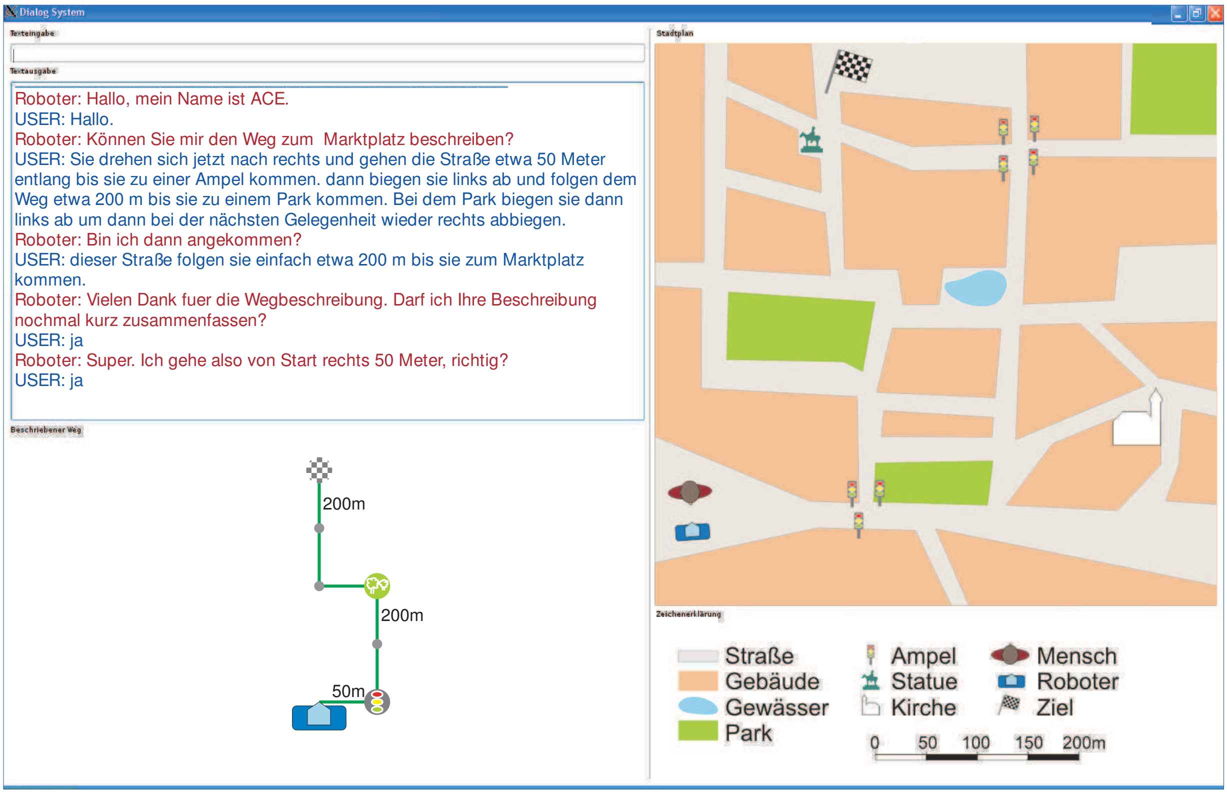


Fig. 3.5: Screenshot of the GUI used during the evaluation experiment.

Each subject interacted with the system twice. Once with the map depicted in Fig. 3.5, and once with the same map rotated by 90° with a different location of robot and human, the succession of the maps was interchanged. Additionally the route belief B_R of the robot, as shown in the bottom left of Fig. 3.5, was only displayed in one of the two runs. A sample dialog of the system with participant H_7 is presented in Table 3.2. After every run the subjects were asked to fill in a questionnaire about system success and user satisfaction. The questionnaire included 14 statements about the system. The statements could be rated with the level of agreement, on a scale between one and five. The statements on the questionnaire are displayed in Table 3.4.

The experimental results are presented and discussed below.

3.5.2 Results

As every subject interacted with the dialog system twice, there are 40 resulting individual interactions, each containing a route description. The average interaction duration was 7 min 17 sec. The average number of outputs from the system was 27 and the average number of inputs from the human subjects was 19.

The experiment is evaluated regarding the guidelines implemented in the dialog system, objective measures of task success, and subjective measures of user satisfaction.

Tab. 3.2: Example dialog between dialog system and human participant H_7.

Robot:	Hello, my name is ACE.
Human:	Hello.
Robot:	Can you tell me, how to get to the market place?
Human:	You have to turn left and follow the street for about 50 meters, until you reach the traffic lights. Then go left and follow the road for approximately 200 meters until you reach a park. Turn left at the park and as soon as possible right again.
Robot:	Am I at my goal then?
Human:	follow the street for about 200 meters to the market place.
Robot:	Thanks for the route description. May I sum up the route briefly?
Human:	yes
Robot:	Great. I have to go right for 50 meters, right?
Human:	yes
Robot:	There I turn left then?
Human:	yes
Robot:	Ok. I have to go straight for 200 meters, right?
Human:	yes
Robot:	Then I have to turn left?
Human:	yes
Robot:	Ok. And then I will turn right?
Human:	yes
Robot:	I have to go straight for 200 meters to the goal, right?
Human:	yes
Robot:	Thank you for your help. Good bye.
Human:	Good bye.

Evaluation of the guidelines

Motivated by DG 1 the dialog phases are implemented as an FSM. This guarantees the possibility to jump back and forth between states GivDir and Conf to correct certain information if necessary. To assess the effectiveness of this structure, the dialogs are analyzed focussing on the active states over the time. Fig. 3.6 shows a box-whisker plot of the durations of the dialog system in the four states Intro, GivDir, Conf, and Concl. The durations for states GivDir and Conf are the cumulative durations, as there may be several switchings between those states. State Intro has a mean duration of 1 min 32 sec with a standard deviation of 1 min 13 sec. GivDir is the central dialog state and has the longest duration as expected, i.e. a mean duration of 6 min 07 sec and a standard deviation of 2 min 37 sec. The mean value of the duration of state Conf is 43 sec while the standard deviation is 25 sec. Finally, Concl has a mean duration of 12 sec with a standard deviation of 5 sec. State GivDir represents the phase in which the crucial information is communicated. The fact that state Intro has the second longest duration with a large standard deviation can be explained easily. During the Introduction phase the human is asked to give directions to a certain goal. The system switches to the next state when the

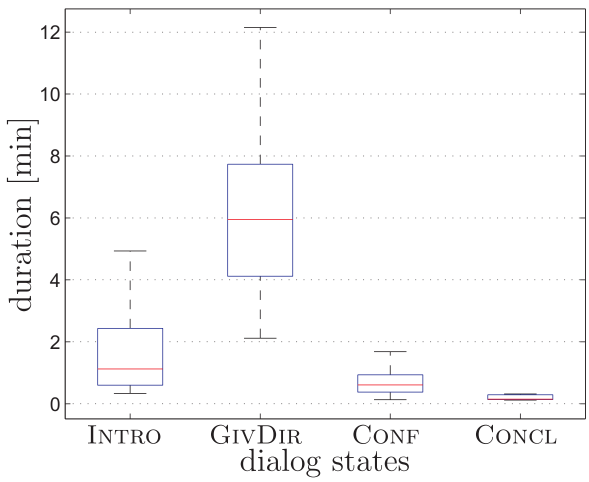


Fig. 3.6: Box-whisker plot of the durations of the dialog states.

human starts giving directions. As argued in [80] some humans first review the whole route and plan the complete route description before beginning to give directions in one big block which leads to long durations of state INTRO; this occurred in 30 % of the dialogs. Others plan the route description piecewise and provide the description sequentially which results in a short duration of INTRO, but prolongs the duration of state GIVDIR; as observed in 70 % of the dialogs in the experiment.

The states and the transitions between them are depicted in Fig. 3.7 for four example dialogs from the experiment. The dialog with participant H_7 is presented in Table 3.2.

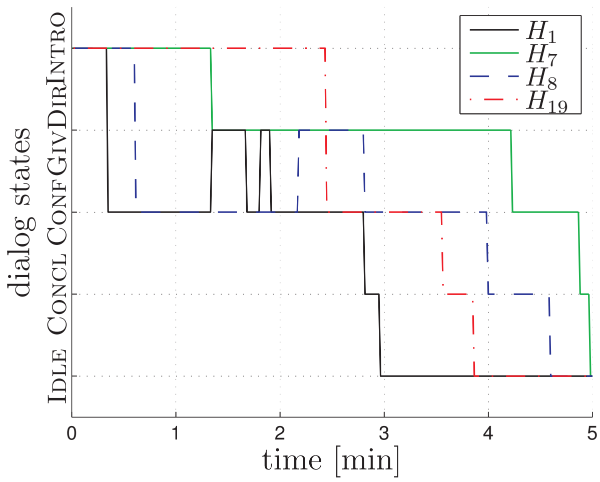


Fig. 3.7: Transitions between system states over the time in four example dialogs from the experiment (participants H_1, H_7, H_8, H_{19}).

As can be seen some dialogs included more than one transition between state GIVDIR and state CONF, where erroneous information was corrected. The mean number of transitions from state CONF back to state GIVDIR is 1.45 with a standard deviation of 1.15. This gives evidence of the necessity of a flexible dialog structure which allows to change back and forth between states and correct faulty information.

DG 2 suggests that the robot asks the human to define the first direction with a pointing gesture and thus solve the coordination problem. Gestures were not included in the simulation presented above, however they were considered in the outdoor experiment with the *ACE* robot [224], as described in Chapter 2. Only in one case the robot was given a faulty direction by gesture. All other gestures were expedient and served to align the personal reference systems of human and robot, so that subsequent route information could be given unambiguously.

As proposed by DG 3 the system extracted route information from the dialogs in the experiment and represented the routes as topological route graphs. The route graphs included directions (DG 4, DG 8), distances (DG 7), and landmarks (DG 5, DG 6). Fig. 3.8 depicts four example route graphs constructed during the experiment. For reference: dialog with participant H_7 is presented in Table 3.2, while the transitions between the states over the time for this participant are depicted in Fig. 3.7. The route graphs vary in detail, especially in the number of nodes, as some subjects gave a more detailed description, naming more landmarks along the way or providing metrical estimates of distances.

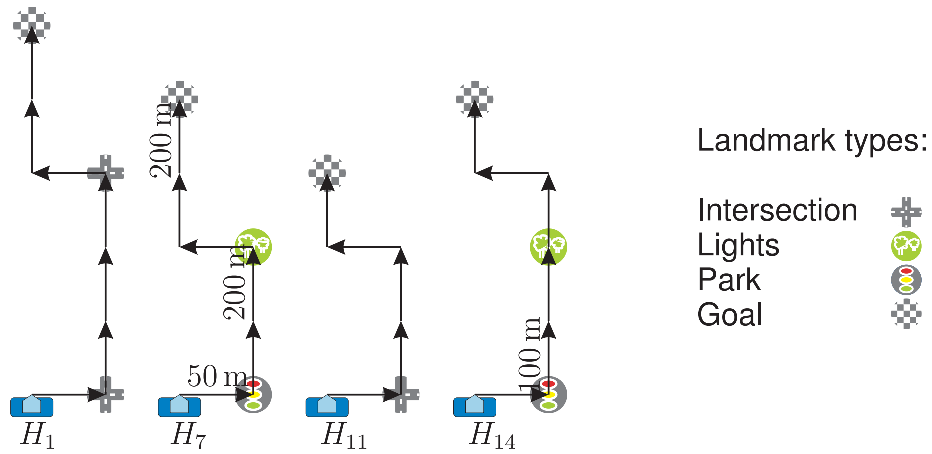


Fig. 3.8: Four routes extracted in the experiment (participants H_1, H_7, H_{11}, H_{14}).

The route belief of the system was depicted in the GUI as proposed by DG 12. This depiction did not result in significant differences of the objective performance, however was rated well subjectively by the humans, as discussed further in the following. The objective results of extracting route information and representing it are discussed below.

Objective Measures

In order to evaluate the performance of the dialog system objectively, it is important to compare it to human performance. To assess the correctness of the route descriptions provided by the subjects during the experiment, all route descriptions were given to other subjects to reconstruct the routes. The subjects were 20 PhD students, 6 female and 14 male, aged between 25 years and 32 years. These subjects were provided with a city map that included the detail map from the GUI and additionally depicted surrounding environment to give more options for interpretation of the route description. A mask to block out parts of the city map and thus restrict the view was used to imitate a movement along the path without further knowledge of the surrounding map. Each subject was asked to reconstruct two route descriptions by drawing a line along a path on the map following the instructions in each route description. The descriptions were rated as correct when the path ended at the goal point which occurred in 24 cases (60 %). Paths that ended along the way to the goal were associated with partially correct route descriptions which happened in one case (2.5 %). In the other cases the route descriptions were rated as incorrect.

The route information extracted by the dialog system was classified in the same way and compared to the results from the route reconstructions by humans. In 11 cases (27.5 %) the extracted routes were correct. The routes were partially correct in 9 cases (22.5 %), i.e. the route was correct, but did not reach all the way to the goal. The number of partially correct routes can be explained by the fact that the system represents each direction information as a route segment delimited by intersections by default if no specific landmarks are extracted which renders some routes too short.

Random guessing results in reaching the goal point in 1.92 %, as there are 52 locations within the same radius as the goal around the robot position. Of those positions 12 are accepted as being located along a route to the goal.

The results of the accuracy of the routes interpreted by humans, extracted by the dialog system, and achieved by random guessing is shown in Table 3.3. The column entitled correct destination lists the accuracy of understanding the route description completely, i.e. reaching the goal. The column entitled partially correct route lists the accuracy of understanding the route description including cases in which the goal point is not reached but the path is correct as far as it reaches. For the cumulated completely and partially correct routes the system performance is $\frac{20}{25} = 80\,\%$ of the human performance which is quite good taking into account that the text recognition does not work perfectly in order to simulate imperfect speech recognition.

Tab. 3.3: Performance given by the accuracy of understanding route descriptions.

	correct destination	partially correct route
human performance	24/40 (60 %)	25/40 (62.5 %)
dialog system	11/40 (27.5 %)	20/40 (50 %)
random guessing	0.77/40 (1.92 %)	9.22/40 (23 %)

The results are in line with the results of other researchers. A direction instruction interpretation algorithm based on probabilistic inference presented by Wei et. al. [197] achieved an accuracy of 85 % of the human performance. However, this system was given keywords extracted by hand from written text as input, and did not process raw text. MacMahon et. al. found that the interpretation of implicit information in route instructions by action inference [112] raises task performance of a virtual agent navigating by human route instructions to 88 % of the human performance, while following purely explicit instructions led to only 41 % of the human performance. Again the system input was “the hand-verified ‘gold-standard’ parse treebank”, not the raw text. Bugmann et. al. presented an instruction based learning system processing speech signals [22] that achieved a recognition rate of 75.9 % compared to human performance. Compared to these studies, the objective results of this work are very good, as the presented system simulates imperfect speech recognition by interpreting raw text input which led to most of the encountered extraction problems. Spell check, or even hand verified keyword input would further improve the system performance.

The recognition accuracy showed no significant differences for the variable of displaying the route belief of the robot in the GUI. Objectively it is not important for information extraction whether the route belief of the robot is displayed or not.

Subjective Measures

The results presented so far are the objective measures of system performance. However in HRI subjective measures such as user satisfaction are important factors for success as well. Therefore a questionnaire has been composed, inspired by established questions for subjectively evaluating dialog systems presented by [196] and [67]. Table 3.4 displays the questionnaire used to evaluate the subjective measures of user satisfaction and subjective assessment of features of the dialog systems. The level of agreement was rated between one (strongly agree) and five (disagree) by the subjects. The mean values and standard deviations of the answers are displayed in Table 3.4 next to the respective questions.

Tab. 3.4: Questions about user satisfaction with subjective assessment.

No.	Statement	μ	σ
1.	The system was easy to use.	2.05	0.93
2.	The system understood the information I entered.	3.25	1.28
3.	I knew at all times what I could enter.	2.78	1.25
4.	The system reacted as I expected it to.	3.20	1.07
5.	The length of the interaction was adequate.	2.40	1.13
6.	The system understood my route description correctly straight away.	3.73	1.38
7.	If not: The system understood my route description correctly after the correction.	2.97	1.48
8.	The structure of the dialog was sensible.	2.13	1.09
9.	I knew from whose point of view I had to describe the route.	1.28	0.64
10.	It is important that the point of view is clear.	1.48	0.85
11.	The system understood landmarks during the route description.	2.38	1.17
12.	It is important that the system understands landmarks.	1.45	0.81
13.	The system understood distance information well.	2.23	1.14
14.	It is important that the system understands distance information.	1.88	0.94

The results of the survey show that the ease of use was rated as good according to statement 1. However, the answers to statements 2, 4, and 6 show that the subjects were not entirely content with what the system understood and how it reacted. In the cases where the users stated that the system did not understand their instructions straight away, a significant improvement (dependent t-test of mean answers: $t(19) = 4.32$, $p < .001$) was established through the correction of information in the CONF phase, as resulting from statements 6 and 7 respectively.

Statements 8 to 14 assess the subjects' views on the usefulness and realization of the implemented guidelines. The results show that the users assessed these rules as reasonable and useful. The implementation of these guidelines was rated well by the subjects, i.e. mean values lie between 1.45 and 2.23.

The answers to statements 6 and 11 about how well the robot understood information depends considerably on the objective results. The mean value of statement 6 for correct routes was 3.00, while for incorrect routes it was 4.16 (independent t-test: $t(18) = -1.47, p = 0.16$). Similarly, the mean values of statement 7 were not significant with 2.17 for correct routes and 3.60 for incorrect routes, (independent t-test: $t(18) = -1.92, p = 0.075$). For statement 11 the mean values of 2.00 for correct and 3.22 for incorrect routes showed a significant difference (independent t-test: $t(18) = -2.38, p = 0.029$). Other differences between ratings of statements for the different objective results were not significant.

All subjects were asked whether the displaying of the robot's route knowledge, as proposed in DG 12, helped during the interaction. This question was answered yes by 14 subjects (70 %). Some gave the explanation that they were not used to interacting with a machine so freely and thus were not always sure whether the system did understand them without the graphical information. Consequently humans interacting with a natural language dialog system seem to prefer some additional feedback modality to be able to evaluate the cognitive capabilities of the system.

Overall, the dialog system was perceived as natural by the human subjects and the implemented dialog guidelines were assessed as useful.

3.6 Discussion

A legitimate approach for robots to close gaps in their route knowledge is to ask humans for directions. Therefore a robot needs a dialog system that can proactively extract route descriptions from human-robot dialogs. There are two main requirements for such a dialog system: the extracted information must be unambiguous, in order to be usable during navigation; the dialog must be natural to humans, in order to facilitate the communication between non-expert human users and a robot.

This chapter presented a dialog system for robots asking humans for directions and proactively extracting the necessary information. As natural language communication is an important skill for robotic systems interacting with non-expert users, principles from linguistics research have been surveyed to identify implications for HRI from human-human communication. Guidelines for robots asking humans for directions have been subsequently derived as policies for human-robot dialogs. A dialog system has been developed with a modular architecture in order to allow for several communication modalities and for extension of discourse topics. The guidelines derived from findings from human-human communication are implemented in the dialog system. These guidelines render the dialog natural for the human partner and enable the system to extract missing route information unambiguously. The dialog system comprises four major components with the dialog manager at its core and an FSM controlling the progress of the dialog.

The dialog system was evaluated in an experiment in which the system interacted with human subjects and extracted route information. Questionnaires about user satisfaction and subjective assessment were evaluated. The results showed that the system is generally natural and easy to use to humans, and that the implemented guidelines were assessed as reasonable by the users. At the same time the system performance of understanding human route directions of 80 % relative to human performance is more than comparable to results from other researchers.

The main contribution of this chapter lies within the successful application of guidelines for human-robot communication derived from human-human communication in a dialog system for robots. The dialog system extracts unambiguous route information and facilitates natural-language dialogs between non-expert human users and robots. The dialog system is designed in a modular way to allow expandability. It can be used with various combinations of input modalities. Other discourse topics can be implemented by adding adequate guidelines for human-robot dialogs.

The dialog system proactively asks humans for missing information and extracts given route descriptions. It does not evaluate extracted information or assess it for plausibility. As route descriptions are abstractions of humans' cognitive representations of spatial relations in the real environments, they can be overly simplified, distorted or even erroneous. Therefore methods are presented for evaluating the reliability of individual route information in Chapter 4 and assessing the plausibility of whole route descriptions in Chapter 5.

4 Probabilistic Models of Route Information

> Probabilistic models for information in route descriptions extracted during human-robot dialogs are presented in this chapter. The presented models for direction and distance information allow robots to assess the reliability and accuracy of the often simplified and distorted information in route descriptions.

4.1 Problem Description and State of the Art

This chapter presents probabilistic models for individual direction and distance information provided by humans in route descriptions. These models can be used by robots to represent information probabilistically and assess the accuracy and reliability of it.

Representational accounts of cognition propose that humans store and represent spatial relations in the real world in so-called cognitive maps, as introduced in [185]. A cognitive map is constructed in a sequence of psychological transformations [40], including change in scale, rotation, and perspective. Route descriptions are verbal abstractions of routes represented in cognitive maps, where a series of processes is involved, such as selection of important elements, temporal structuring, and selection of reference frames [49]. Therefore a route description involves two transformations of spatial relations, i.e. from the real world into a cognitive map [40] and from the cognitive map into a verbal description [49]. These transformations are not mathematical mappings, but cognitive processes, and therefore prone to simplifications, inaccuracy, and even errors. Thus, information in route descriptions is not exact, but noisy.

Robots, constructing a route belief from human route descriptions, need information models that account for the uncertainty of route information. Route information that can be modeled probabilistically are directions, quantitative distances, i.e. metric distance values, and qualitative distances, such as distance ranges described by spatial deictics. Based on a probabilistic representation of information, robots can establish the reliability and accuracy of the acquired information. This can be used as one of several measures when merging route information from different sources. A process of comparing different route descriptions and building a probabilistic representation by merging the route information is presented in Chapter 5.

Researchers are trying to model the cognitive processes involved in mentally representing spatial relations [57, 132, 187]. These cognitive models have even been applied to technical systems [91]. Probabilistic models of cognition reflect the degrees of belief [29, 62]. There is a large body of literature presenting models of cognitive representations of spatial relations, while there is very little research done on modeling the verbal abstractions of these cognitive representations in the form of route descriptions. However, for the application of robots extracting spatial information from HRI, probabilistic models of information in route descriptions are needed. These probabilistic models enable robots to construct a route belief for navigation based on a probabilistic interpretation of the information.

The most essential information embedded in route descriptions are directions between decision points, i.e. landmarks. Direction information may be erroneous because of an incomplete [120] or oversimplified [25, 107] cognitive spatial representation, or because of difficulties in conveying the spatial representation as for example in the case of left-right confusion [15]. Distortions of cognitive directions have been described by [180, 194] which analyze continuous angular directions, rather than discrete distinct directions used in route descriptions. Route direction structure diagrams [149] have been introduced as tools for the structural analysis of direction information; they depict aspects of route descriptions such as directions over decision points. Subjective concepts of directions in route descriptions have been researched in [83] as well. However, to date there are no works on modeling the reliability or soundness of direction information in route descriptions.

Other essential information given in route descriptions are distances between decision points. Researchers have modeled cognitive distances, i.e. the distance represented in an individual's cognitive map. There are different approaches to how the distances were given in such studies, e.g. route distances vs. flight distances, and to which functions should model the relationship between cognitive and real distance. Linear functions have been used to model cognitive distances within cities [36] and between cities [26]. However, the results of those works and of other researchers, e.g. [107], show that there is a tendency to overestimate small and great distances, while underestimating distances in medium range which points out that the linear function is not the most appropriate model. Furthermore, the linear function has been found to be inappropriate, because the constant term does not approximate to zero as would be expected [18]. Cognitive distances were also modeled by power functions [16, 18], as the power function is generally accepted as the psychophysical law [173]. The cognitive distance depends on factors such as emotional involvement [16], familiarity with a route [35], the location relative to the city center [18, 55, 102], and even vegetation along a route [157]. The perceptual relationship between visual distance perception and real distance has been modeled as well [54], based on a mathematically defined metric for visual space [108]. The perceived distance has been found to depend on the real distance and a finite limit of perceived distance as a rational function, where both the numerator and the denominator are first order polynomials [54]. Several external and individual influencing factors to visual perception of distances have been identified [47, 144]. A probabilistic model for visual space depending on the topological properties of the environment based on Bayesian inference is discussed in [205].

Researchers have found that the remembered distance is generally smaller than the perceived distance [14, 125, 127]. In all experiments cognitive distances or perceived distances were compared to the real distances. The cognitive distance, however, is not the same as the distance humans give in route descriptions which is a much coarser estimate of the real distance, often provided as salient travelling distance or time values. Again, there are no research results on the conceptual relationship between real distances and distances estimated during while giving route descriptions.

This chapter presents probabilistic models of information in route descriptions, in particular of directions and quantitative as well as qualitative distances. These models can be applied to route information extracted by the dialog system presented in Chapter 3. They allow robots to reason about individual segments within route descriptions, while a system for reasoning about whole route descriptions is introduced in Chapter 5. Inspiration for such probabilistic models for route information is found in probabilistic robot mapping [182], in Markov chains [115], and Bayesian inference [12]. These probabilistic models are meant for technical systems to serve as tools for assessing the accuracy and dependability of individual information in route descriptions extracted from dialogs with humans. The probabilistic modeling of direction and distance information in route descriptions in itself goes beyond current spatial cognition research that focuses on models of cognitive directions and distances. Furthermore, the presented models find specific applications in robotics, as means to assess the accuracy of given route information.

This chapter is structured as follows. Section 4.2 presents probabilistic models for direction and distance information from a theoretic point of view. In Section 4.3 survey data is analyzed and numerical models for direction, quantitative, and qualitative distance information are extracted. An evaluation is presented in Section 4.4, where an exemplary route is represented probabilistically. Section 4.5 discusses the presented models.

4.2 Models for Direction and Distance Information

Route information, such as directions and distances, are imprecise representations of relations in the real environment and can even be erroneous. Therefore a robot that has to build a route belief from route directions given by humans requires a model of the accuracy and reliability of such information. A common way to consider such uncertainty is by modeling it probabilistically. In this chapter theoretical approaches to probabilistic models for directions and distances are investigated. Fig. 4.1 schematically shows the process of probabilistically modelling route information extracted from human-robot dialogs.

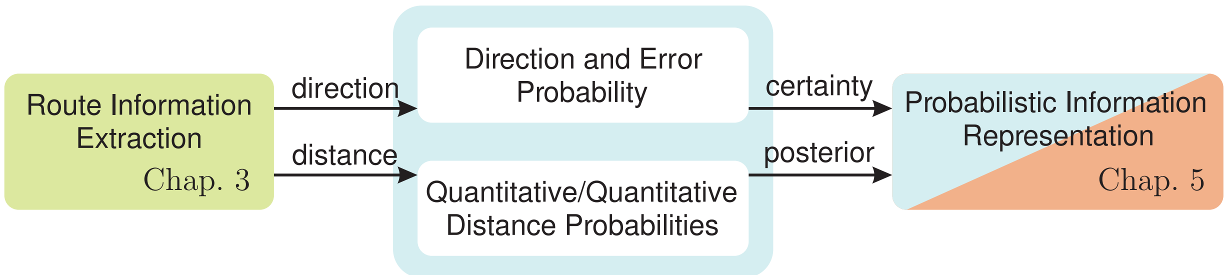


Fig. 4.1: Overview over the process of modeling route information probabilistically.

4.2.1 Certainty Value of Direction Information

The aim of this subsection is to enable a technical system to assign certainty values to each direction information in a route description, as an assessment of their reliability. Such certainty values have to comprise all factors that can possibly influence the reliability of direction information. The reliability of direction information in a route description clearly depends on the route segment k it is given in. The probability for a certain direction of route segment k depends on the current direction, the direction of the previous route segment and the direction of the previous route up to route segment $k-1$. Additionally the probability that route segment k is erroneous is regarded in calculating the reliability of a direction information.

On the way to obtain certainty values for direction information a suitable representation of directions has to be chosen. At first, all direction information is transformed into global directions relative to the orientation in the point of origin. The succession of directions in a route description can be regarded as a descriptive process which has a distinct direction state at each route segment k. The discrete set of direction states

$$S = \{\boldsymbol{s}_1,\, \boldsymbol{s}_2,\, \boldsymbol{s}_3,\, \boldsymbol{s}_4\} = \{\boldsymbol{s}_{\text{left}},\, \boldsymbol{s}_{\text{straight}},\, \boldsymbol{s}_{\text{right}},\, \boldsymbol{s}_{\text{back}}\}\ , \tag{4.1}$$

includes the four orthogonal directions that are represented as unit vectors, e.g. $\boldsymbol{s}_1 = \begin{pmatrix}1 & 0 & 0 & 0\end{pmatrix}^T$. The observations are the directions at each route segment k in a route description which are composed of fractions of distinct direction states. A discrete global direction $\boldsymbol{x}_k$ in the route description is represented as direction state vector

$$\boldsymbol{x}_k = \begin{pmatrix}x_1 & x_2 & x_3 & x_4\end{pmatrix}^T = \begin{pmatrix}x_{\text{left}} & x_{\text{straight}} & x_{\text{right}} & x_{\text{back}}\end{pmatrix}^T\ , \tag{4.2}$$

with fractions $0 \leq x_i \leq 1$ of discrete directions and $\sum\limits_{i=1}^{4} x_i = 1$. The description of a direction observation as a state vector allows for the application of a Markov chain inspired approach for the calculation of the direction probabilities. Furthermore, the general direction model takes into account not only the four orthogonal directions, but non-orthogonal directions such as *'sharp left'*, *'veer right'*, or numerical angles, as well.

Direction Probability

A direction probability model is derived from transition probabilities between the distinct directions $\boldsymbol{x}_k$. The state transition probabilities are given by

$$p_{ij} = P\left(\boldsymbol{x}_k = \boldsymbol{s}_j | \boldsymbol{x}_{k-1} = \boldsymbol{s}_i\right) \tag{4.3}$$

and have the properties $p_{ij} \geq 0$ and $\sum\limits_{j=1}^{4} p_{ij} = 1$. Furthermore the probabilistic description is time independent, as it is truncated to the current and the predecessor state:

$$p_{ij} = P\left(\boldsymbol{x}_k = \boldsymbol{s}_j | \boldsymbol{x}_{k-1} = \boldsymbol{s}_i,\, \boldsymbol{x}_{k-2} = \boldsymbol{s}_h,\, \ldots\right) = P\left(\boldsymbol{x}_k = \boldsymbol{s}_j | \boldsymbol{x}_{k-1} = \boldsymbol{s}_i\right) \tag{4.4}$$

A transition matrix $\boldsymbol{P}$ comprises all state transition probabilities p_{ij}, such that

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & \cdots & p_{14} \\ \vdots & \ddots & \vdots \\ p_{41} & \cdots & p_{44} \end{bmatrix}\ . \tag{4.5}$$

The transition matrix $\boldsymbol{P}$ applies to route segments $k > 1$. Initial state probabilities π_i for route segment $k = 1$, are combined in an initial state probability vector

$$\boldsymbol{\pi} = P(\boldsymbol{x}_i) = \begin{pmatrix}\pi_1 & \pi_2 & \pi_3 & \pi_4\end{pmatrix}^T , \tag{4.6}$$

with the properties $\pi_i \geq 0$ and $\sum_{i=1}^{4} \pi_i = 1$.

The initial state probabilities $\boldsymbol{\pi}$ and the state transition probabilities $\boldsymbol{P}$ are used to calculate the direction probability $P_{D,k}$ of each route segment k which depends on the global direction $\boldsymbol{x}_{k-1}$ of the previous route segment and the global direction $\boldsymbol{v}_{k-1}$ of the overall previous route up to the previous route segment. An overal direction vector $\boldsymbol{v}_k$ is of unit length $\sum_{i=1}^{4} v_i = 1$ and includes entries that are fractions of discrete directions $0 \leq v_i \leq 1$. The direction of the previous route segment is considered by multiplying the global direction vector $\boldsymbol{x}_{k-1}$ and the probability matrix. The prediction vector $\tilde{\boldsymbol{x}}_k$ of the direction of $\boldsymbol{x}_k$ which is independent of the route before segment $k-1$, is given by a Markov chain of length one, as

$$\tilde{\boldsymbol{x}}_k = \boldsymbol{x}_{k-1}\,\boldsymbol{P}\,. \tag{4.7}$$

Additionally, the prediction $\tilde{\boldsymbol{v}}_k$ of direction $\boldsymbol{x}_k$ depending on the overall previous route

$$\tilde{\boldsymbol{v}}_k = \boldsymbol{v}_{k-1}\,\boldsymbol{P}\,, \tag{4.8}$$

is considered, because it can be expected that a global overall direction is followed in the long-run. The predictions $\tilde{\boldsymbol{x}}_k$ in (4.7) and $\tilde{\boldsymbol{v}}_k$ in (4.8) are combined in a direction prediction $\hat{\boldsymbol{x}}_k$ by an entry-wise multiplication in a Hadamard product (denoted by $\tilde{\boldsymbol{x}}_k \circ \tilde{\boldsymbol{v}}_k$) and a normalization by division by the dot product. Prediction $\hat{\boldsymbol{x}}_k$ considers both the direction of the previous segment and the direction of the overall previous route:

$$\hat{\boldsymbol{x}}_k = \frac{\tilde{\boldsymbol{x}}_k \circ \tilde{\boldsymbol{v}}_k}{\tilde{\boldsymbol{x}}_k^T \tilde{\boldsymbol{v}}_k} \tag{4.9}$$

The probability $P_{D,k}$ of the direction $\boldsymbol{x}_k$ at route segment k is gained by multiplying the predictions by the actual direction, such that

$$P_{D,k} = \begin{cases} \boldsymbol{\pi}^T\,\boldsymbol{x}_k, & \text{if } k = 1 \\ \hat{\boldsymbol{x}}_k^T\,\boldsymbol{x}_k, & \text{otherwise} \end{cases} . \tag{4.10}$$

The direction probability $P_{D,k}$ takes into account the current direction at k, the previous direction at $k-1$, and the overall direction of the previous route up to $k-1$. It does not yet consider the possibility that route segment k might be erroneous at route segment k.

Error Probability

To assess the reliability of an individual direction information, not only the direction probability $P_{D,k}$ is important, but also the probability $P_{E,k}$ that a route segment is erroneous. Erroneous direction information in a route description at route segment k, affects the subsequent direction information in such a way that the following segments are erroneous as well. Therefore the error probability $P_{E,k}$ of an individual direction information in a route description is modeled by a cumulative distribution, as

$$P_{E,k} = \int_{-\infty}^{k} p_E(x) \mathrm{d}x \,, \tag{4.11}$$

where $p_E(k)$ is the probability density function (pdf) of the first occurence of an erroneous route segments. A cumulative distribution function conforms to the boundary conditions that $P_{E,k}$ increases with the route segment k. The error probability is used as a heuristic to weight the direction probability and result in a certainty value for each route segment k. $P_{E,k}$ reflects the fact that information at the beginning of a route description is in general more reliable than information that is farther away in the description.

Certainty Value

The direction probability P_D and the error probability P_E are mixed by a weighted average, in analogy to sensor models in probabilistic robot mapping [182]. The resulting certainty value is defined by the weight vector

$$\boldsymbol{w} = \begin{bmatrix} w_D & w_E \end{bmatrix}^T \,, \tag{4.12}$$

with weight w_D for the direction probability and weight w_E for the error probability, where $w_D + w_E = 1$. The certainty value results to

$$c_k = \boldsymbol{w}^T \begin{bmatrix} P_{D,k} \\ 1 - P_{E,k} \end{bmatrix} \,. \tag{4.13}$$

The certainty value c_k is assigned to the direction information of each segment in the route belief of a robot and provides an assessment of the reliability of the information. In the case that the direction plausibility $P_{D,k}$ or $P_{D,k-1}$ equals zero, the weight w_D is set to one and subsequently the certainty value c_k is assigned the value zero, as the whole route is assumed to be faulty after an erroneous route segment. In this way each segment in a route description is assigned an individual certainty value which reflects the reliability of that particular direction information.

4.2.2 Posterior Probability of Distance Information

Distance information provided in route descriptions is usually a rounded value of the cognitive representation of the real distance. Therefore a probabilistic model for such distance estimates is needed for robots that are using route descriptions from humans as a basis for global navigation. Probabilistic distance models allow robots to assess during navigation whether a location in the real environment corresponds to the respective description.

Posterior Probability

Bayesian inference [10] is used to determine the probability that a location in the environment at the real distance d_{real} corresponds to an estimate d_{est} in a route description. This probability is given by the posterior probability which calculates as

$$P(d_{\text{real}}|d_{\text{est}}) = \frac{P(d_{\text{est}}|d_{\text{real}})\,P(d_{\text{real}})}{P(d_{\text{est}})}\,, \tag{4.14}$$

with the prior probability $P(d_{\text{real}})$, the marginal probability $P(d_{\text{est}})$, and the conditional probability $P(d_{\text{est}}|d_{\text{real}})$.

A robot can assess whether a distance in the real environment is likely to correspond to a given distance estimate, by evaluating the relative posterior. If the relative posterior $P_{\text{rel}}(d_{\text{real}}|d_{\text{est}})$ is above a certain threshold the real distance may well correspond to the distance described. The condition is given by

$$P_{\text{rel}}(d_{\text{real}}|d_{\text{est}}) = \frac{P(d_{\text{real}}|d_{\text{est}})}{\max\left(P(d_{\text{real}}|d_{\text{est}})\right)} \geq \tau_{P_{\text{rel}}}\,. \tag{4.15}$$

Additionally, the posterior of a real distance can be compared to other candidate real distances, in this way determining the most probable real distance given a distance estimate in a route description.

Prior Probability

The real distance is distributed uniformly, as every real distance value is as likely as another, therefore the prior probability depends only on the distance range r_{real}, such that

$$P(d_{\text{real}}) = \frac{1}{r_{\text{real}}}\,. \tag{4.16}$$

The distance range r_{real} results from the boundaries of the considered environment.

Marginal Probability

Distance and time estimations in route descriptions do not reflect the real distance or the cognitive distance, but are usually roughly rounded estimates of these. They are not uniformly distributed, but rather the frequency increases for some round and salient values or personal preferences. The marginal probability $P(d_{\text{est}})$ takes into account that the probabilities for estimations are not distributed uniformly, and balances higher frequencies of certain salient estimates in the posterior probabilities.

Conditional Probability

The conditional probability $P(d_{\text{est}}|d_{\text{real}})$ constitutes the likelihood that an estimation value d_{est} given during a route description refers to a certain real distance d_{real}. The conditional probability is the crucial component for calculating the posterior probability.

The models for direction and distance information presented above provide robots with the means of assessing route information for plausibility and reliability. In order to obtain metric values for the certainty values and posterior probabilities, numerical values are required in the models. These specific numerical values are obtained by collecting and analyzing route information data.

4.3 Data Analysis

Theoretical approaches to equip robots with probabilistic models of route information have been discussed above. Numerical values for the theoretical models are derived from survey data. In this way technical systems are provided with specific models for direction, quantitative and qualitative distance information. This work focuses on robots operating in urban environments on sidewalks and in pedestrian areas. Therefore the models derived in the following are intended and valid for walking distances in such environments, while for other scales of environments the models may differ and must be extracted from appropriate data.

4.3.1 Direction Information

A specific model for the reliability of relative directions in route descriptions is needed by robotic systems in order to assess the plausibility of this information in route descriptions. Such a model is provided by the certainty value c_k in (4.13) which depends on the direction probability $P_{D,k}$, and the error probability $P_{E,k}$. These probabilities are calculated from the analysis of data from real route descriptions.

Route Description Data

Route description data was collected in the city of Munich in 2009 [222]. In total 48 passers-by were randomly selected and asked to give directions to well known goals within walking distance. All dialogs were recorded with a voice recorder and transcribed to text. In 5 cases the persons asked did not know the destination and could not give a route description, resulting in 43 analyzable dialogs. In order to analyze the route descriptions with regard to direction accuracy, the route directions were verified on a city map, and erroneous route segments were marked. For a more representative amount of data, route descriptions collected in Trier [61] were evaluated additionally. There are 101 route descriptions in the dataset from Trier, in which erroneous route segments were marked as well. There is no significant difference between the lengths of the route descriptions, i.e. the numbers of route segments (t-test, $p = 0.4339$) and therefore the route description data sets are assumed to belong to the same distribution. Thus the route descriptions from Trier and Munich are evaluated collectively. The combined data set contains 144 route descriptions. The 92 correct route descriptions and 24 erroneous descriptions result in an overall error ratio of $\frac{24}{116} \approx 21\%$. The number of segments in the route descriptions ranges from one to seven with a mean value of 2.811 and a standard deviation of 1.075.

Direction Probability

The direction probability $P_{D,k}$ in (4.10) is calculated from the current and previous global direction state vectors and the state transition matrix $\boldsymbol{P}$, or the initial probability vector $\boldsymbol{\pi}$. The numerical variables that need to be derived to model the direction probability are the state transition probabilities p_{ij} and the initial state probabilities π_i. Therefore the data is analyzed by transforming the extracted distance information in all route descriptions into global directions and analyzing the transition frequencies between the individual states, or by analyzing the frequencies of the initial directions in the route descriptions.

Analysis of the directions of segment $k = 1$ of the route descriptions provides numerical values for the initial state probability vector $\boldsymbol{\pi}$ in (4.6), as

$$\boldsymbol{\pi} = \begin{bmatrix} 0.09 & 0.77 & 0.10 & 0.04 \end{bmatrix} . \tag{4.17}$$

The transition probabilities might look surprising at first glance, as opposed to an expectation of having equal probabilities for all states, the transition probabilities are varying. It makes sense however, as the data have been collected in real experiments where not all orthogonal direction options were available which explains the relatively high probability π_2 for the direction *'straight'*. The initial probabilities π_1 for *'left'* and π_3 for *'right'* are found to be approximately the same, as expected from unbiased data.

The transition matrix (4.5) results from the frequencies of the transitions in the data to

$$\boldsymbol{P} = \begin{bmatrix} 0.56 & 0.18 & 0 & 0.26 \\ 0.25 & 0.50 & 0.25 & 0 \\ 0 & 0.12 & 0.58 & 0.30 \\ 0.28 & 0 & 0.26 & 0.46 \end{bmatrix} . \tag{4.18}$$

As expected the main diagonal, i.e. the probabilities for transitions between the same global directions have the highest values of $p_{ii} \approx 0.5$, while transitions to states representing orthogonal directions are $p_{i(i\pm1)} \approx 0.25$, and the transition probabilities to states of reverse directions are $p_{i(i\pm2)} = 0$.

The initial state probabilities π_i in (4.17) and the state transition probabilities p_{ij} in (4.18) are depicted in Fig. 4.2 as a digraph of transition probabilities between direction states in a route description.

It is noted that the transition probabilities are not symmetrical. Although, a large number of route descriptions with a number different combinations of start and end points has been evaluated to identify these transition probabilities, a bias towards certain directions cannot be excluded. However, the numerical initial and transition probabilities provide a sound model to compute the direction probability $P_{D,k}$ in (4.10).

Error Probability

In order to find a specific model $P_{E,k}$ for the error probability, the erroneous route descriptions are analyzed further. An important question is in which route segment k the first error occurs. For this analysis only route descriptions with $k \leq 5$ are used, as there are too few longer route descriptions to draw significant conclusions from. The relative error

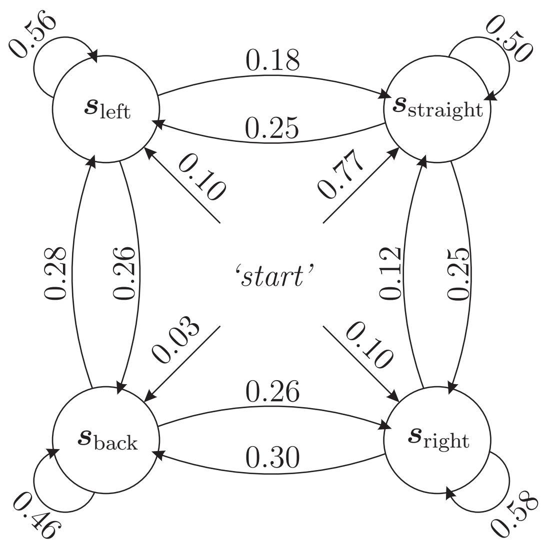


Fig. 4.2: Digraph of transition probabilities between direction states.

frequency $f_{E,k}$, i.e. the ratio of errors at route segment k, is depicted in Fig. 4.3 marked by squares, over the number of route segments k. The relative frequency $f_{E,k}$ represents the first occurrence of an error in the route descriptions in the data set.

Erroneous direction information in a route description affects the subsequent direction information in such a way that the directions in the following segments are erroneous as well. Therefore the cumulative relative error frequency $F_{E,k} = \sum_{i=1}^{k} f_{E,k}$ represents the overall ratio of errors in each route segment. The cumulative frequency F_k is marked by circles in Fig. 4.3 over the number of route segments k.

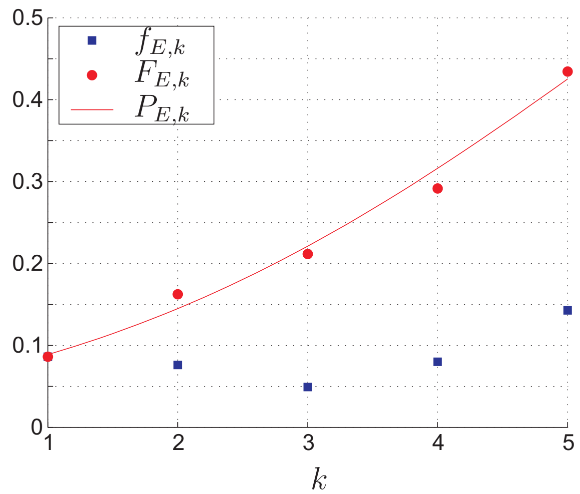


Fig. 4.3: Relative frequency $f_{E,k}$, relative cumulative frequency $F_{E,k}$, and fitted cumulative probability distribution $P_{E,k}$ of an error at route segment k.

The cumulative error frequency $F_{E,k}$ increases with the number of route segments k. As there is not enough route description data with more than five route segments the shape of the function for $k > 5$ is unknown, but is expected to converge to one, as it is assumed that the longer a route description is the likelier it becomes that there is an error in it. Therefore the error probability distribution is modeled as a normal cumulative probability distribution, such that

$$P_{E,k} = \frac{1}{\sigma_E\sqrt{2\pi}} \int\limits_{-\infty}^{k} e^{\frac{-(x-\mu_E)^2}{2\sigma_E^2}} \mathrm{d}x \,. \tag{4.19}$$

The method of least squares minimizes the vertical quadratic error between the cumulative relative frequency and the fitted normal function, with

$$\min_{\mu_E,\sigma_E} \sum_{k=1}^{5} \left(F_{E,k} - P_E(k|\mu_E,\sigma_E)\right)^2 \ , \ \text{s.t. } (4.19)\,. \tag{4.20}$$

In this way the values $\mu_E = 5.7$ and $\sigma_E = 3.3$ are identified. The error model $P_{E,k}$ is depicted in Fig. 4.3 as a line.

Certainty Value

The certainty value c_k in (4.13) is fully described by the probabilities P_D and P_E and the weight vector $\boldsymbol{w}$. The weights w_D and w_E are assigned heuristically for the probabilistic direction information model c_k. The weight w_E influences the offset and weight w_D influences the variance of the certainty value c_k. The certainty values for all direction information in the route description data set is calculated to demonstrate the procedure and outcome. Fig. 4.4 shows the resulting probabilities $P_{D,k}$ as dotted lines, $1 - P_{E,k}$ as a dashed line, and the resulting certainty values c_k as solid lines.

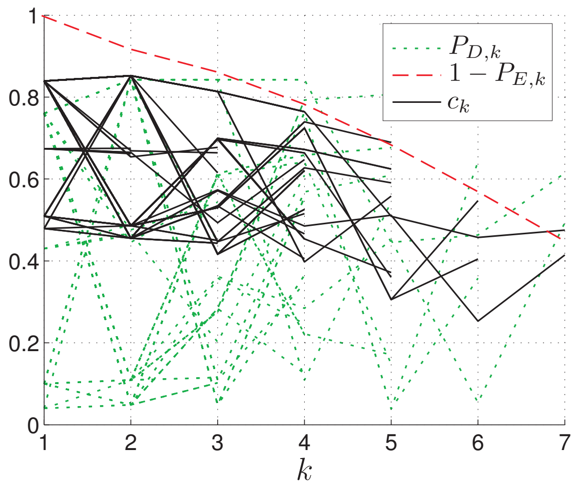


Fig. 4.4: Overview of direction and error probabilities and resulting certainty values for all direction information in the data set of route descriptions. Direction probability $P_{D,k}$ in dotted lines, error probability $1 - P_{E,k}$ in dashed line, and the resulting certainty values c_k in solid lines, with $\boldsymbol{w} = \begin{bmatrix}0.5 & 0.5\end{bmatrix}$.

The probabilities $P_{D,k}$ and $1 - P_{E,k}$ and the certainty value c_k decrease with a growing number of route segments k, as the probabilities that either the route description is at its end or an error occurs increases with every route segment. The certainty values c_k provide a robot with an assessment of the reliability of the respective direction information.

4.3.2 Quantitative Distance Information

Robots navigating based on global route information extracted from HRI, need specific models of the accuracy of distance information given in route descriptions. Such models of quantitative distance estimates allow them to assess whether a location in the real environment can be allocated to a location mentioned in a route description by a distance estimate. Quantitative distance information in route descriptions is typically provided as a numerical travelling distance d_{est} or time t_{est} estimate. The functions and numerical values of the posterior probabilities $P(d_{\mathrm{real}}|d_{\mathrm{est}})$ and $P(d_{\mathrm{real}}|t_{\mathrm{est}})$, the prior probability $P(d_{\mathrm{real}})$, the conditional probabilities $P(d_{\mathrm{est}}|d_{\mathrm{real}})$ and $P(t_{\mathrm{est}}|d_{\mathrm{real}})$, and the marginal probabilities $P(d_{\mathrm{est}})$ and $P(t_{\mathrm{est}})$ are identified by analyzing collected survey data.

Quantitative Distance Estimation Data

As distance estimates are not included in every route direction, distance estimation data was collected in a separate survey by means of asking 110 randomly selected passers-by for the way and the remaining distance to three different well-known landmarks in the center of Munich. The participants were all asked at different locations to estimate how far it was to one of the three landmarks which was not visible but within walking distance. In order to record the distance estimation in a route description and not the more precise cognitive distance, the subjects were not told that the data was for a scientific survey until after they gave their estimations, as not to influence their estimation efforts. Interestingly after they were told that the data was intended for a scientific study, some participants wanted to modify their distance estimate. This supports the assumption that a distance estimate given in a route description differs from the respective cognitive distance. Each participant gave either an estimation of the remaining distance, usually in [m], or the remaining time, usually in [min], or in some cases both. In total the data consists of 71 distance estimates and 65 time estimates. The real distances d_{real} are rounded to an accuracy of 10 m, and range from 100 m to 1000 m.

Fig. 4.5 on the left depicts the real distances d_{real} over the estimated distances d_{est} of the data collected in the survey. A dashed line marks the correspondence of real and estimated distance $d_{\mathrm{est}} = d_{\mathrm{real}}$. The real distances d_{real} over the estimated times t_{est} are shown in Fig. 4.5 on the right. A dashed line marks the correspondence of the real distance with the estimated time multiplied by an average walking velocity as $t_{\mathrm{est}}\, v_{\mathrm{walk}} = d_{\mathrm{real}}$, with the walking velocity $v_{\mathrm{walk}} = 4.5\,\frac{\mathrm{km}}{\mathrm{h}}$.

Generally it is noted that for short real distances d_{real} the estimates tend to be lower, whereas for longer distances the estimates tend to be higher than the real values.

Effects of Personal Factors on the Distance Estimation Accuracy: The data of 60 of the subjects includes each subject's gender, age, self-assessment of the estimation, and an explanation of the given self-assessment. The effects of the personal factors gender, age, and self-assessment on the estimation accuracy are evaluated. For that purpose,

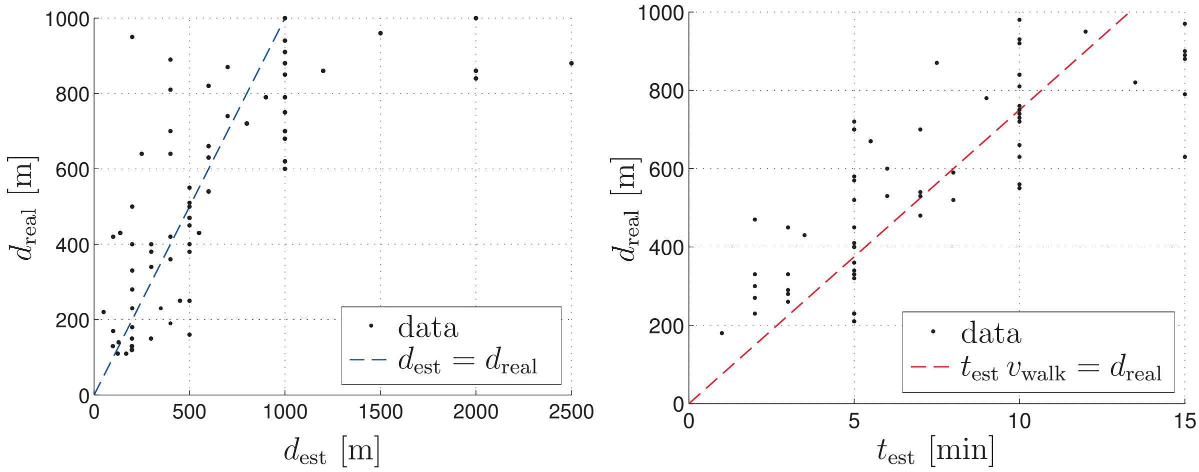


Fig. 4.5: Estimation data of traveling distances on the left and traveling times on the right. The dashed lines indicate the match of estimated and real data.

relative constant and absolute errors of the distance and time estimation data are compared statistically with t-tests and analyses of variances (ANOVAs), respectively. The full results are given in Appendix B.1, while condensed results are presented here.

Both relative constant and absolute errors for the time data are significantly lower than for the distance data which suggests that time estimates are generally more reliable than distance estimates. The personal variable gender does not show significant differences between data from female and male participants for the relative errors of distance and time estimates respectively, however female participants showed a preference for giving time estimates, while males preferred to give distance estimates. The differences between three different age groups of the relative errors of the data were also not significant. Similarly, the variable self-assessment did not provide significant relative error differences between three different groups of self-assessment, except for the relative constant distance estimation error which arose from the fact that participants who assessed their estimate as *'bad'* tended to underestimate distances more often than other participants.

The fact that differences of the relative errors are not significant for the variables gender, age, and self-assessment, with only one exception, shows that those personal variables have no important influence on the accuracy of quantitative walking distance or time estimation within route descriptions. Therefore these factors do not have to be taken into account by a robot when searching for and selecting a person whom to ask for directions. Furthermore these personal factors are of no importance when modeling the relation between real distance and estimated distance or time.

Prior Probability

The real distance data d_{real} was limited to a range of 100 m to 1000 m in an urban environment, resulting in a distance range $r_{\text{real}} = 900$ m for which this model is valid. Thus, the prior probability in (4.16) is

$$P(d_{\text{real}}) = \frac{1}{900}\,. \tag{4.21}$$

Marginal Probabilities

The marginal probabilities $P(d_{\mathrm{est}})$ and $P(t_{\mathrm{est}})$ consider a non-uniform distribution of the frequency of the estimates, i.e. a preference of humans to provide salient and rounded values as distance estimates. Histograms of the frequencies $f_{d_{\mathrm{est}}}$ and $f_{t_{\mathrm{est}}}$ of the data are presented in Fig. 4.6, for distance estimates on the left, and time estimates on the right. The histograms display higher frequencies for salient distances, e.g. 500 m and 1000 m, and salient times, e.g. 5 min and 10 min, slightly raised frequencies between those salient values, and an overall decrease in the frequency with an increasing real distance.

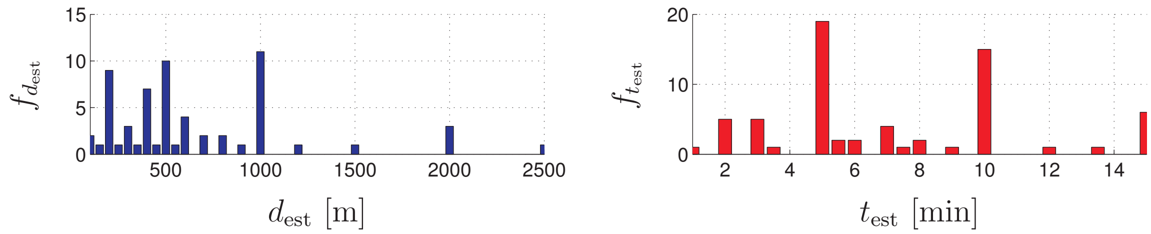


Fig. 4.6: Histograms of quantitative distance and time estimates.

The marginal probabilities can be determined by a lookup table of the relative frequencies or alternatively by a fitted function. In both cases it has to be considered that the experimental data has gaps at some estimation values which have to be interpolated.

Conditional Probabilities

Conditional probabilities reflect the likelihood that a distance d_{est} or time t_{est} estimate is provided in a route description given a certain real distance d_{real} in the environment. It is assumed that these probabilities are modeled by normal functions, where the mean values and standard deviations are functions of the real distance. The conditional probability for distance estimates is given by

$$P(d_{\mathrm{est}}|d_{\mathrm{real}}) = P(d_{\mathrm{est}}|\mu_d,\, \sigma_d) = \frac{1}{\sigma_d\sqrt{2\pi}}\, e^{\frac{-(d_{\mathrm{est}}-\mu_d)^2}{2\sigma_d^2}}\,, \tag{4.22}$$

where μ_d and σ_d are functions of d_{real}. Analogously the conditional probability for time estimates is modeled by

$$P(t_{\mathrm{est}}|d_{\mathrm{real}}) = P(t_{\mathrm{est}}|\mu_t,\, \sigma_t) = \frac{1}{\sigma_t\sqrt{2\pi}}\, e^{\frac{-(d_{\mathrm{est}}-\mu_t)^2}{2\sigma_t^2}}\,, \tag{4.23}$$

where μ_t and σ_t are functions of d_{real}. The functions for the mean values μ_d and μ_t represent the conceptual relations between the real distance d_{real} and the estimated distance d_{est} and estimated time t_{est} respectively. An appropriate function for describing these variables is chosen in the following.

The data distribution, as shown in Fig. 4.5, is consistent with the data distributions for perceived distances within visual space [54] and with data distributions for cognitive distances [16]. In the literature different types of functions are found modeling perceived and remembered distance over real distances. Therefore the two most common functions, a rational function and a power function are compared and subsequently the more appropriate function is used to describe μ_d, μ_t, σ_d, and σ_t.

Approaches applying rational functions and power functions found in the literature are both applied to the collected data and compared by their residuals. A rational function with first order polynomials as numerator and denominator similar to the model for visually perceived distances [54] is given by

$$g_{\mathrm{rat}}(d_{\mathrm{real}}) = \frac{d_{\mathrm{real}}}{\alpha_1\, d_{\mathrm{real}} + \alpha_2}\,. \tag{4.24}$$

A power function in line with results for cognitive distances [16, 107] is written as

$$g_{\mathrm{pow}}(d_{\mathrm{real}}) = \beta_1 {d_{\mathrm{real}}}^{\beta_2}\,. \tag{4.25}$$

The two alternative functions are fitted to the data sets for distance and time estimations by the method of least squares.

The functions $g_{\mathrm{rat}}(d_{\mathrm{real}})$ in (4.24) and $g_{\mathrm{pow}}(d_{\mathrm{real}})$ in (4.25) are compared by the sum of squared residuals of the data points $\left(d_{\mathrm{real},i} \quad d_{\mathrm{est},i}\right)$ and $\left(d_{\mathrm{real},i} \quad t_{\mathrm{est},i}\right)$ and the modeled values. The sum of squared residuals are

$$\begin{aligned} r^{\mathrm{rat}}_{y_{\mathrm{est}}} &= \sum_{i=1}^{n} \left(y_{\mathrm{est},i} - g_{\mathrm{rat}}(d_{\mathrm{real},i}, \alpha_1, \alpha_2)\right)^2\,, \\ r^{\mathrm{pow}}_{y_{\mathrm{est}}} &= \sum_{i=1}^{n} \left(y_{\mathrm{est},i} - g_{\mathrm{pow}}(d_{\mathrm{real},i}, \beta_1, \beta_2)\right)^2\,, \end{aligned} \tag{4.26}$$

with the data type $y_{\mathrm{est}} = d_{\mathrm{est}}$ for distance or $y_{\mathrm{est}} = t_{\mathrm{est}}$ for time estimates. They are compared for data with stepwise increased maximum real distances. Fig. 4.7 compares $r^{\mathrm{rat}}_{y_{\mathrm{est}}}$ to $r^{\mathrm{pow}}_{y_{\mathrm{est}}}$ over the maximum real distance $\max(d_{\mathrm{real}})$, as solid lines for $g_{\mathrm{rat}}(d_{\mathrm{real}})$ and as dashed lines for $g_{\mathrm{pow}}(d_{\mathrm{real}})$, for distance estimates on the left and time estimates on the right.

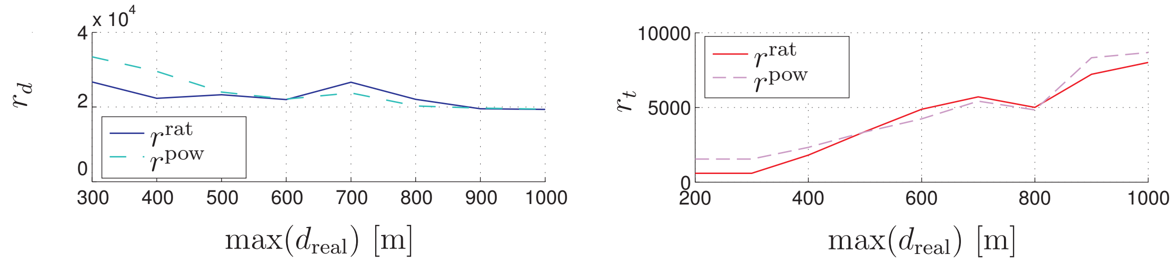


Fig. 4.7: Sums of squared residuals $r^{\mathrm{rat}}_{y_{\mathrm{est}}}$ and $r^{\mathrm{pow}}_{y_{\mathrm{est}}}$ over the maximum real distance compare the goodness of fit of functions $g_{\mathrm{rat}}(d_{\mathrm{real}})$ and $g_{\mathrm{pow}}(d_{\mathrm{real}})$ for distance estimates on the left and time estimates on the right.

The sum of squared residuals is smaller for $g_{\mathrm{rat}}(d_{\mathrm{real}})$ for a small real distance maximum under 500 m and again for a big maximum distance at 1000 m both for the distance and time estimation data, while the sum of squared residuals is smaller for $g_{\mathrm{pow}}(d_{\mathrm{real}})$ for a real distance maximum in the medium range. This gives evidence that the rational function $g_{\text{rat}}(d_{\text{real}})$ provides a better model for the estimated distance and time information. Therefore the rational function $g_{\text{rat}}(d_{\text{real}})$ in (4.24) is chosen to model the characteristics μ_d, σ_d, μ_t, and σ_t of the conditional probabilities. The numerical values of the coefficients α_1 and α_2 obtained by the method of least squares are presented in Table 4.1.

Tab. 4.1: Identified coefficients of function $g_{\text{rat}}(d_{\text{real}})$ for distance and time estimations applied to describe μ_d, σ_d, μ_t, and σ_t of the respective conditional probabilities.

	distances		times	
	α_1	α_2	α_1	α_2
μ	0.0009	0.5157	0.0008	0.0057
σ	0.0069	1.3622	0.0079	0.0258

The left of Fig. 4.8 depicts the fitted functions for μ_d as a solid line and $\mu_d \pm \sigma_d$ as dotted lines for distance estimates. On the right of Fig. 4.8 the fitted function μ_t is shown as a solid line and $\mu_t \pm \sigma_t$ are depicted as dotted lines for the time estimates.

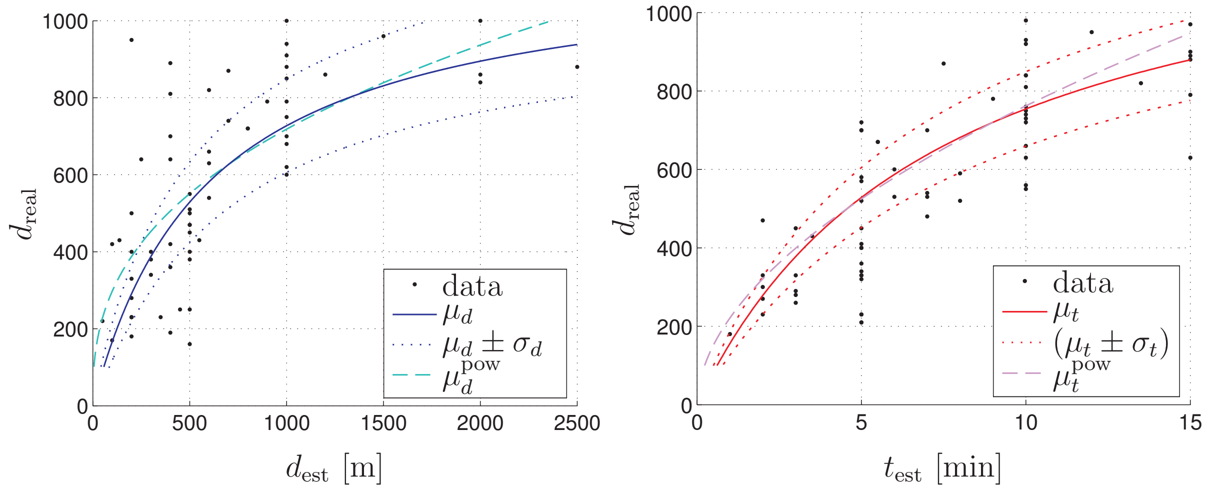


Fig. 4.8: The fitted function $g_{\text{rat}}(d_{\text{real}})$ of the mean values μ of the conditionals as a solid line, and of the intervals of one standard deviation $\mu \pm \sigma$ in dashed lines for distance estimates on the left and for time estimates on the right. The alternative results for μ using the power function $g_{\text{pow}}(d_{\text{real}})$ are depicted in light dotted lines.

Additionally the mean values fitted by the power function $g_{\text{pow}}(d_{\text{real}})$ are depicted in Fig. 4.8 as dashed lines. It is confirmed graphically that the rational function fits the data better than the power function especially for short and big distances.

In summary, the variables μ_d, μ_t, σ_d and σ_t of the conditional probabilities $P(d_{\text{est}}|d_{\text{real}})$ in (4.22) and $P(t_{\text{est}}|d_{\text{real}})$ in (4.23) are described by rational functions $g_{\text{rat}}(d_{\text{real}})$ in (4.24) with the coefficients given in Table 4.1. The presented models are valid for real distances between 100 m and 1000 m.

Posterior Probabilities

The posterior probabilities $P(d_{\mathrm{real}}|d_{\mathrm{est}})$ and $P(d_{\mathrm{real}}|t_{\mathrm{est}})$ in (4.14) are computed using the prior (4.21), marginal , and conditional (4.22), (4.23) probabilities. The resulting density distribution for $P(d_{\mathrm{real}}|d_{\mathrm{est}})$ is depicted in Fig. 4.9 on the left for the valid range of 100 m to 1000 m for d_{real} and 100 m to 2500 m for d_{est}. The density distribution $P(d_{\mathrm{real}}|t_{\mathrm{est}})$ is depicted on the right of Fig. 4.9 for the valid range of 100 m to 1000 m for d_{real} and 1 min to 15 min for t_{est}. The overall curved shapes of the posterior distributions are evoked by

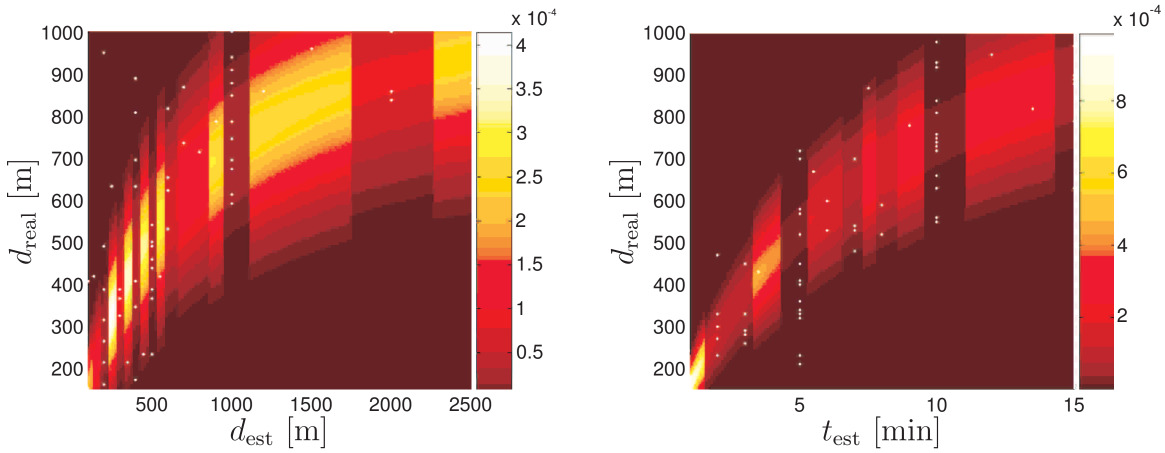


Fig. 4.9: The distributions of posterior probabilities $P(d_{\mathrm{real}}|d_{\mathrm{est}})$ for distance estimates on the left and $P(d_{\mathrm{real}}|t_{\mathrm{est}})$ for time estimates on the right.

the rational function in the conditional probabilities which model the tendencies to over- and underestimate certain distance ranges. The bands with different widths are caused by the marginal probability distributions accounting for the tendency of humans to give salient values as estimations.

When a robotic agent follows route descriptions from humans, the posterior probability $P(d_{\mathrm{real}}|d_{\mathrm{est}})$ or $P(d_{\mathrm{real}}|t_{\mathrm{est}})$ gives a probability value for a reached real distance d_{real} to be the sought-after distance, given a travelling distance estimate d_{est} or time estimate t_{est}, respectively. If the relative posterior probability in (4.15) is above a certain predefined threshold, e.g. $P_{\mathrm{rel}}(d_{\mathrm{real}}|d_{\mathrm{est}}) \geq 0.1$, or is high compared to the relative posteriors of other candidate real distances, the real distance can be assumed to be the distance described by the estimate in the route description.

4.3.3 Qualitative Distance Information

Distance information in route descriptions can be specified not only explicitly as metrical distance or time estimates, but also implicitly by using spatial deictics. Frequently used deictics referring to distances are *'here'*, *'close'*, and *'far'*. These deictics describe distance ranges as opposed to specific distance values. These distance ranges are stochastic and are modeled probabilistically as posterior probabilities $P(d_{\mathrm{real}}|\,\textit{'here'})$, $P(d_{\mathrm{real}}|\,\textit{'close'})$, and $P(d_{\mathrm{real}}|\,\textit{'far'})$ in (4.14) to give robots models of the locations in the real environments given one of these deictics in the route descriptions. The posterior probabilities depend on the prior probability $P(d_{\text{real}})$, the conditional probabilities $P(\textit{'here'}|d_{\text{real}})$, $P(\textit{'here'}|d_{\text{real}})$, and $P(\textit{'far'}|d_{\text{real}})$, and the marginal probabilities $P(\textit{'here'})$, $P(\textit{'close'})$, and $P(\textit{'far'})$.

Qualitative Distance Estimation Data

An experiment was conducted with 36 PhD students of engineering as participants. There were 4 female and 32 male participants aged between 24 years and 37 years, with an average age of 27 years. The participants were given a questionnaire showing a map of the city center of Munich. A start point was marked in the center of the map and 44 locations marked by boxes were randomly assigned around it , as shown in Fig. C.5 in Appendix C.2. The box farthest away from the start point corresponded to a real distance of 1130 m in the corresponding real environment. The participants were all familiar with the area around the start point. They were asked to label the boxes as either *'here'*, *'close'*, or *'far'*, if they could clearly associate them with one of these deictics.

Contextual Dependence: Before modeling qualitative distances, a question that needs to be answered is whether the deictics *'here'*, *'close'*, and *'far'* are contextually independent, i.e. whether they are used to refer to relative distances from a starting point independently of the considered environment. An experiment was conducted to discover whether or not the deictics *'here'*, *'close'*, and *'far'* describe different distance ranges in different environments. In the experiment the borders between pairs of adjacent deictics, are analyzed depending on the given environment. The experiment is based on a questionnaire, as presented in Appendix C.2, where the participants had to mark the borders between *'here'* and *'close'*, and between *'close'* and *'far'* in three different maps of different environments, i.e. an abstract environment, an urban environment, and a large scale environment.

The results of the evaluation of contextual dependence are presented in detail in Appendix B.2; here the results are summarized. A mixed linear model is fitted to the data and analyzed statistically. The variable map is statistically significant with a strong effect both on the borders between *'here'* and *'close'* and between *'close'* and *'far'*. The variable age is significant for the border between *'here'* and *'close'*, but has only a weak effect. Thus it is argued that the environment has to be taken into account when modeling qualitative distance information, while effects of personal variables can be neglected. Therefore the models for the spatial deictics *'here'*, *'close'*, and *'far'*, presented in the following hold only for the considered urban environment.

Prior Probability

The distance data d_{real} was limited to a range of 0 m to 1130 m in an urban environment, resulting in a distance range of $r_{\text{real}} = 1130\,\text{m}$ for which this model is valid. Thus, the prior probability in (4.16) is

$$P(d_{\text{real}}) = \frac{1}{1130}\,. \tag{4.27}$$

Marginal Probabilities

The marginal probabilities $P(\textit{`here'})$, $P(\textit{`close'})$, and $P(\textit{`far'})$ denote the likeliness that the respective deictic is used to describe a distance in a route description. These probabilities are inferred from the frequency of the use of the deictics *'here'*, *'close'*, and *'far'* in real route descriptions, i.e. from the data set described in Section 4.3.1. In the data set the deictic *'here'* is given 29 times, while *'close'* and *'far'* are used twice each. The marginal probabilities correspond to the frequencies of the data, resulting in

$$\begin{aligned} P(\textit{`here'}) &= \frac{29}{33}, \\ P(\textit{`close'}) &= \frac{2}{33}, \\ P(\textit{`far'}) &= \frac{2}{33}. \end{aligned} \tag{4.28}$$

Conditional Probabilities

The results are evaluated separately for the deictics *'here'*, *'close'*, and *'far'*, and truncated functions are used to model the probability distributions in the following.

The frequencies f_h of the physical distances between the start point and all locations that were marked as *'here'* are shown in the histogram on the top of Fig. 4.10. The start point is always referred to as *'here'*. The farther the distance is from the start point, the less likely it is referred to as *'here'*. To model the probability density function for the deictic *'here'* a truncated exponential function is fitted to the data by the method of least squares. The resulting distribution is

$$P(\textit{`here'}|d_{\mathrm{real}}) = \begin{cases} \dfrac{\frac{1}{\mu_{\mathrm{h}}} e^{\frac{d_{\mathrm{real}}}{\mu_{\mathrm{h}}}}}{1 - e^{\frac{1130}{\mu_{\mathrm{h}}}}}, & \text{if } 0 \leq d_{\mathrm{real}} \leq 1130 \\ 0, & \text{otherwise} \end{cases}, \tag{4.29}$$

with $\mu_{\mathrm{h}} = 148.67$ identified by the method of least squares. The conditional probability $P(\textit{`here'}|d_{\mathrm{real}})$ is shown on the bottom of Fig. 4.10.

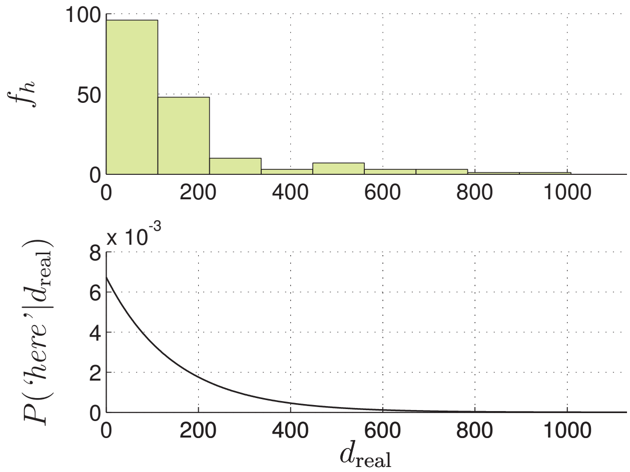


Fig. 4.10: Histogram of the data for *'here'* on top. The resulting conditional probability distribution modeled by a truncated exponential function on the bottom.

The frequency f_c of the data of deictic *'close'* is depicted in Fig. 4.11 as a histogram. It resembles a lognormal distribution, therefore a truncated lognormal function is fitted to the data by the method of least squares, such that

$$P(\text{'}close\text{'}|d_{\text{real}}) = \begin{cases} \dfrac{\frac{1}{d_{\text{real}}} e^{\frac{-(\ln(d_{\text{real}})-\mu_c)^2}{2\sigma_c^2}}}{\int\limits_0^{1130} \frac{1}{x} e^{\frac{-(\ln(d_{\text{real}})-\mu_c)^2}{2\sigma_c^2}} dx - 1}, & \text{if } 0 \leq d_{\text{real}} \leq 1130 \\ 0, & \text{otherwise} \end{cases}, \tag{4.30}$$

where $\mu_c = 5.81$ and $\sigma_c = 0.51$ are identified. The probability density function is shown on the bottom of Fig. 4.11.

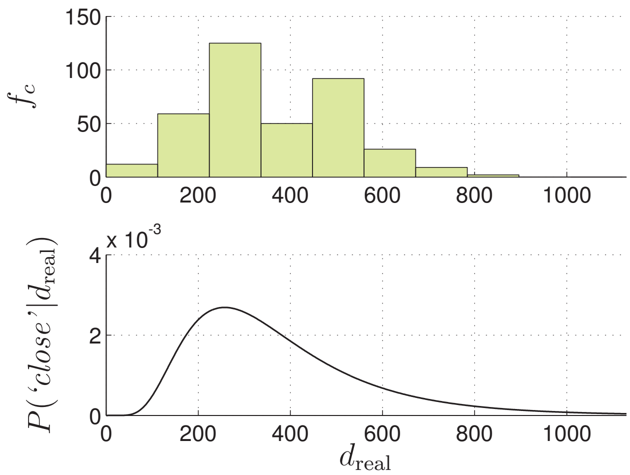


Fig. 4.11: Histogram of the data for *'close'* on the top, and the resulting conditional probability distribution modeled as a truncated lognormal function on the bottom.

The frequency f_f for the data of the deictic *'far'* over the relative distance d_{real} is shown on the top of Fig. 4.12. A right truncated Weibull function approximates the data, as

$$P(\text{'far'}|d_{\mathrm{real}}) = \begin{cases} \dfrac{\alpha_{\mathrm{f}}\, \beta_{\mathrm{f}}\, {d_{\mathrm{real}}}^{\beta_{\mathrm{f}}-1}\, e^{-\alpha_{\mathrm{f}}\, {d_{\mathrm{real}}}^{\beta_{\mathrm{f}}}}}{1 - e^{-\alpha_{\mathrm{f}}\, 1130^{\beta_{\mathrm{f}}}}}, & \text{if } 0 \le d_{\mathrm{real}} \le 1130 \\ 0, & \text{otherwise} \end{cases} . \tag{4.31}$$

The variables $\alpha_{\mathrm{f}} = 1510.07$ and $\beta_{\mathrm{f}} = 3.82$ are identified by the method of least squares. The pdf $P(\text{'far'}|d_{\mathrm{real}})$ is depicted in Fig. 4.12 on the bottom.

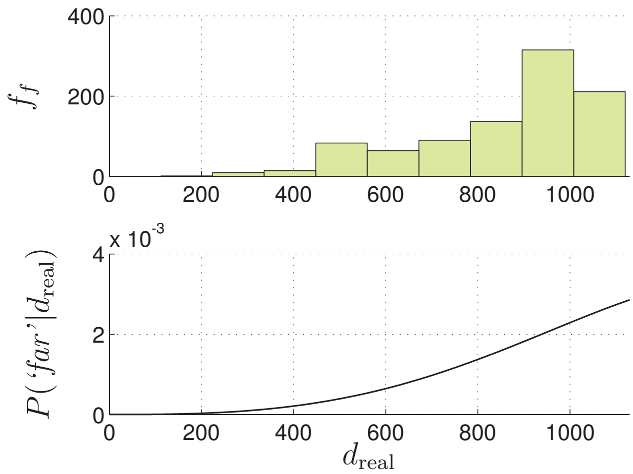


Fig. 4.12: The histogram of the data for *'far'* on the top. The resulting conditional probability distribution modeled as a right truncated Weibull function on the bottom.

The identified conditional probability distributions $P(\text{'here'}|d_{\mathrm{real}})$, $P(\text{'close'}|d_{\mathrm{real}})$, and $P(\text{'far'}|d_{\mathrm{real}})$ are crucial for calculating the posterior probabilities of real distances given route information in the form of deictics *'here'*, *'close'*, or *'far'*.

Posterior Probabilities

The posterior probabilities $P(d_{\mathrm{real}}|\text{'here'})$, $P(d_{\mathrm{real}}|\text{'close'})$, and $P(d_{\mathrm{real}}|\text{'far'})$ in (4.14) are computed using the modeled prior probabilities in (4.27), marginal probabilities in (4.28), and conditional probabilities in (4.29), (4.30), and (4.31). The three resulting posterior pdf's for the deictics *'here'*, *'close'*, and *'far'* are depicted in Fig. 4.13.

The resulting posterior probability density functions $P(d_{\mathrm{real}}|\text{'here'})$, $P(d_{\mathrm{real}}|\text{'close'})$, and $P(d_{\mathrm{real}}|\text{'far'})$ model the distance reaches of the spatial deictics *'here'*, *'close'*, and *'far'*, in the real environment. They can be used to assess whether a feature in the real environment corresponds to the information given in a route description by assigning a threshold for the relative posteriors in (4.15).

Analogously, to quantitative distance estimates, qualitative distances can be given in the time domain, e.g. *'now'* or *'soon'*. However, this occurs too rarely to model posteriors. In the case it does occur, the expressions can be simply mapped to their spatial equivalents and posterior probabilities are computed as described above.

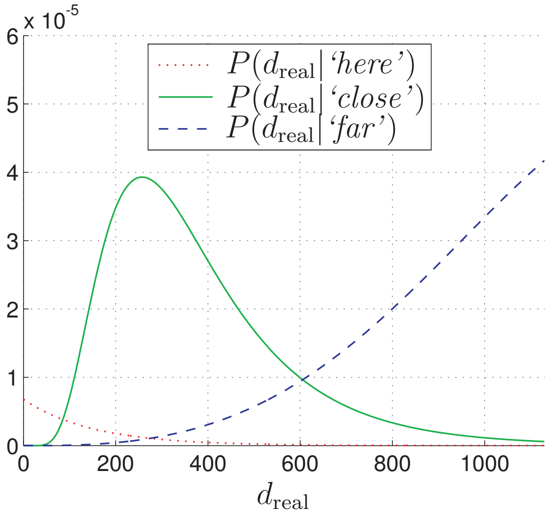


Fig. 4.13: Posterior distance probabilities for the qualitative distance expressions *'here'*, *'close'*, and *'far'* over the real distance in an urban environment.

4.4 Evaluation

The soundness of the assignment of certainty values to direction information, and of posterior distributions to distance information is demonstrated by applying the models presented above to a sample route description.

An example route between two locations within the city center of Munich is depicted in Fig. 4.14 on the right. An according route description can be given as *'Go straight for 500 meters and turn right. Veer left after 3 minutes. Then the market is close by.'* The direction and distance information given in this route description is visualized by arrows of proportional directions and lengths in the center of Fig. 4.14.

The direction and distance information given in the route description is represented probabilistically using the models identified above. The direction probability density distribution $P_{D,k}$ in (4.10) is calculated and certainty values c_k are assigned to the direction information according to (4.13), with $\boldsymbol{w} = \begin{bmatrix}0.5 & 0.5\end{bmatrix}$. The posterior distributions for distance information in (4.14) for the quantitative distance estimate, the quantitative time estimate, and the qualitative distance estimate in the route description are computed. The accuracy of the information in the route description is visualized on the right of Fig. 4.14 in the reference system of a robot. The route segments are depicted as proportional arrows of different lightness according to the corresponding certainty value (a darker color signifies a higher certainty value). The direction probability distribution and the posterior distance distributions are multiplied and depicted as point clouds of likely locations of the next decision point relative to the last.

An overview of the values of relevant data in the evaluation is presented in Table 4.2. The data presented for each route segment k are the corresponding real distances *dreal*, certainty values c_k, probabilities of the distance information $P_{D,k}$, relative distance probabilities $P_{D,\text{rel}} = \frac{P_{D,\text{rel},k}}{\max\left(P_{D,\text{rel},k}\right)}$, posteriors of the direction $P(d_{\text{real}}|d_{\text{est}})$, and relative direction posteriors $P_{\text{rel}}(d_{\text{real}}|d_{\text{est}})$ in (4.15).

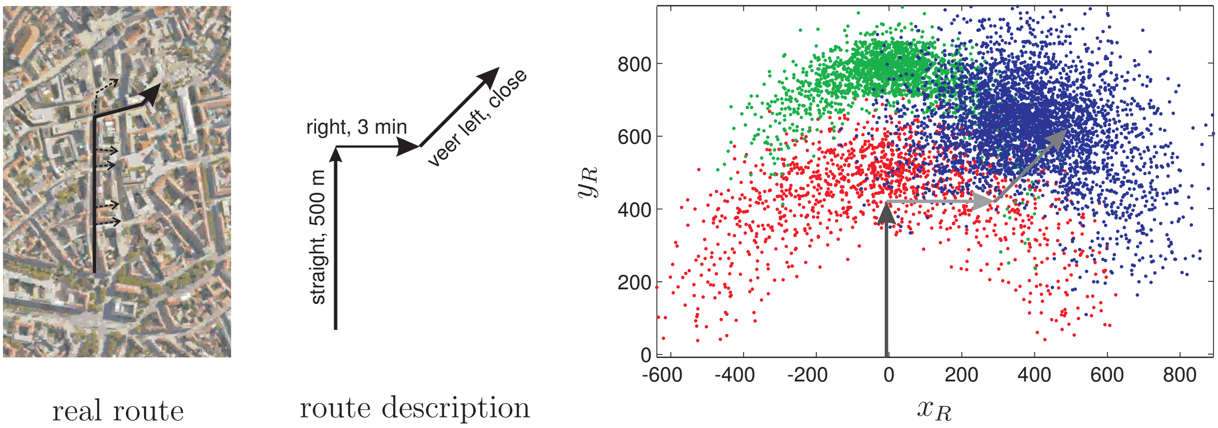


Fig. 4.14: Modeled accuracy of information in an exemplary route description. Left: route in the real environment, alternative branch-offs from the first route segment are marked as dashed lines; center: direction and distance information from a corresponding route description; right: modeled accuracy of route information, where the lightness of an arrow reflects the certainty value of the direction information it represents, the direction probabilities multiplied with the distance posterior distributions are illustrated by overlaid point clouds for each node.

Tab. 4.2: Numerical values of relevant data of route information in the example.

k	d_{real}	c_k	$P_{D,k}$	$P_{D,\text{rel}}$	$P(d_{\text{real}}\|d_{\text{est}})$	$P_{\text{rel}}(d_{\text{real}}\|d_{\text{est}})$
1	414 m	0.8392	0.7600	1.0000	$3.1112 \cdot 10^{-5}$	0.5436
2	134 m	0.4554	0.0485	0.0576	$2.8848 \cdot 10^{-8}$	0.0003
3	110 m	0.6167	0.4489	0.7309	$1.3697 \cdot 10^{-4}$	0.2467

The table presents numerical values for the modeled accuracy depicted in Fig. 4.14. Interestingly, the relative distance posterior probability at route segment $k = 2$ is very low, i.e. $P_{\text{rel}}(d_{\text{real}}|d_{\text{est}}) < 0.1$, which gives evidence to the fact that the corresponding time information was a poor estimate of the real walking time.

The relative posteriors $P_{\text{rel}}(d_{\text{real}}|d_{\text{est}})$ of the alternative branch-offs from the first route segment, as depicted in Fig. 4.14 on the left as arrows with dashed lines, are compared in Table 4.3. As can be seen there are real distances in the physical environment that achieve lower and some that achieve higher relative posterior probabilities than the correct distance $d_{\text{real}} = 414$ m, given the estimated distance of 500 m.

The results show that the presented models provide reasonable assessments of the accuracy of individual information given in route descriptions and are thus suitable as information models. They are not sufficient as the only means to assess the accuracy and reliability of route information for robot navigation and global path planning. Therefore the next chapter presents a system for reasoning about whole route descriptions, assessing them for plausibility, and combining plausible information in the route belief.

Tab. 4.3: Relative posteriors of alternative real distances of $k = 1$, given $d_{\text{est}} = 500\,\text{m}$.

d_{real} [m]	$P_{\text{rel}}(d_{\text{real}} \mid d_{\text{est}})$
130	0.0006
174	0.0030
281	0.0585
325	0.1465
414	0.5436
470	0.8522
612	0.7264

4.5 Discussion

Direction or distance information in route descriptions is usually a simplified, distorted, or even erroneous representation of relations in the real environment. Therefore robots need to take these uncertainties into account when building a route belief for navigation and global path planning based on such information.

This chapter presented probabilistic models for individual direction and distance information occurring in route descriptions. These models facilitate the assessment of the accuracy and reliability of uncertain route information. Probabilistic models are provided for route information extracted for example by the dialog system presented in Chapter 3. The modeled information includes directions, quantitative distances, i.e. distance and time estimates, and qualitative distances in the form of deictic distance descriptions. Direction information is modeled in this chapter as a descriptive process. The model simultaneously takes into account the conditional probability of the direction information given the direction of the previous route and the likelihood of an error occurring during the length of the route description. The direction and error probabilities are combined as a weighted average analogously to sensor models, resulting in a certainty value for distance information. Distance information in route descriptions is modeled by posterior probabilities which allow to assess whether a real distance is the sought-after distance given an estimate. The posterior probability distribution takes into account people's tendency to overestimate some distance ranges while underestimating others, as well as a preference for naming salient values when estimating distances in route descriptions. The presented specific models with their numerical values are valid for walking distances within urban environments. A generalization of the models would be possible by collecting an even larger set of data and including the effects of contextual factors.

The probabilistic models for directions and quantitative and qualitative distances have been evaluated by applying them to an example route description and calculating certainty values and distance posteriors for all information. The evaluation showed that the developed models serve well for assessing reliability and accuracy of information in route descriptions, but are not sufficient as the only means to assess route information. Therefore, the next chapter presents a system that compares route descriptions, and builds a route belief from plausible information.

The main contributions of this chapter are the derived probabilistic models for individual information in route descriptions. As related research focussed only on modeling cognitive representations of directions and distances which are different from information given in route descriptions, the presented probabilistic models are novel to spatial cognition and spatial computation. Additionally, these models are well suited to robotics, as they provide means to assess the reliability and accuracy of route information.

Route information represented probabilistically by the presented route information models reflects the reliability and accuracy of certain given data and can be used as a measure of certainty when merging route descriptions from different sources. These models can only be applied to evaluate individual direction and distance information within route descriptions, but not whole route descriptions. A system for comparing different whole route descriptions and building a probabilistic representation by merging the plausible descriptions is presented in Chapter 5.

5 Simultaneous Reasoning and Mapping

This chapter proposes a system for *Simultaneous Reasoning and Mapping* (*SRAM*). It assesses whole route descriptions extracted from human-robot dialogs for plausibility, represents them internally as route graphs, and adds plausible information to a route belief. The proposed system allows for detecting and correcting gaps, excess, rotations, or errors in route graphs.

5.1 Problem Description and State of the Art

While the approach of extracting missing route knowledge by asking humans for directions is fast and adaptable, information extracted in this way can be simplified, distorted, or erroneous. While Chapter 4 presents probabilistic models to assess direction and distance information of individual route segments for accuracy and reliability, this chapter aims at providing a robot with a procedure to represent and reason about whole route descriptions extracted from human-robot dialogs. This is necessary as different route descriptions have different degrees of refinement or can even be erroneous.

Researchers are working on methodologies for representing and reasoning about spatial information. On the basis of the cognitive map [185] as the cognitive representation of spatial relations in environments, Kuipers [91, 92] has developed a computational theory, the SSH. A probabilistic approach of combining metric and topological map knowledge is presented by Thrun et al. [183], who joined the ideas of the SSH with classical robot mapping. Werner et al. [199] propose modeling navigational knowledge as route graphs. In an extended approach [164] route descriptions in HRI are represented as Voronoi-based route graphs including metric information. Brosset et al. [20] model human route descriptions based on locations and actions. Recently, qualitative spatial reasoning [32, 148] has been applied to robot navigation and path planning. Researchers have started developing robots that can extract and store spatial information about their environments or even route descriptions from human-robot interaction, among others [101, 119, 171, 197]. However, none of these robots reason about the spatial information provided by humans.

The main contribution of this chapter is a general method for *Simultaneous Reasoning and Mapping* (*SRAM*) that processes route descriptions extracted from human-robot dialogs. The framework includes a dialog system, as presented in Chapter 3, which interfaces with human partners and extracts route information from human-robot dialogs. Each extracted direction and distance information is represented probabilistically, as described in Chapter 4. In this way certainty values for directions and posteriors for distances are calculated. The *SRAM* system represents extracted route descriptions as topological route graphs including certainty values, reasons about these routes based on existing route belief, inquires about conflicting information, and integrates plausible route information into the route belief of the robot. As this extracted route information is not necessarily correct or complete, the system performs pattern matching and vector similarity assessment to evaluate whether or not the information is plausible. If necessary the system inquires about conflicting route information, such as missing or excess route segments, or rotations. Finally, plausible information is included in the route belief of the robot. In this way, a reasonable representation of the route in the environment is obtained and forms the basis for global navigation and path planning.

This chapter is structured as follows. The main types of possible differences between route descriptions are reviewed in Section 5.2. As a general procedure for representing extracted route information and reasoning about it, the *SRAM* system is introduced in Section 5.3. An evaluation of the system is presented in Section 5.4. Finally, the chapter is concluded in Section 5.5.

5.2 Types of Differences between Route Descriptions

There are several possible types of errors or differences between route descriptions that can occur. The most common differences between routes are depicted in Fig. 5.1.

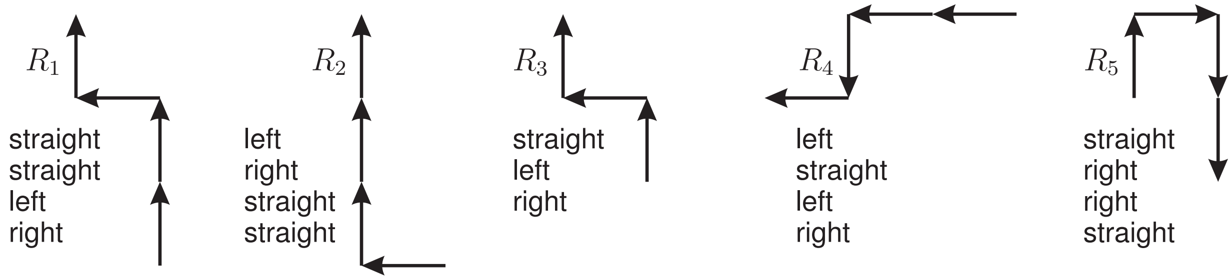


Fig. 5.1: Examples of different route descriptions that may represent the same route with some differences or errors.

The most common types of differences between route descriptions are:

- Two routes can be different from one another but still lead to the same goal, as for example routes R_1 and R_2.

- A routes can have missing route segments which may hint to a different degree of refinement compared to another route description, as in the case of example route R_1 compared to R_3.

- Reversely a route can have excess route segments, as in the case of example route R_3 compared to R_1.

- Routes may have some different segments, resulting in rotations, as shown by example routes R_1 and R_4.

- Finally, two routes can be completely different from one another, such as route R_5 compared to R_1; concluding that at least one of them is not plausible.

The different types of errors or differences between routes have to be assessed for plausibility by different methods. Different routes that lead to the same goal can be identified by assessing the similarity of the vectors spanned between the start and the end node of the routes. Excess or missing route segments, as well as rotation, caused by differing route segments, can be found by pattern matching. Both route similarity assessment and pattern matching are adequate to classify completely different routes as implausible. A system is needed that binds the methods for plausibility assessment together and builds a plausible route belief. This system is introduced in the following section.

5.3 Simultaneous Reasoning and Mapping

An approach to solve the problem of reasoning about route information extracted from HRI and simultaneously building a route belief as an internal representation of the plausible route information is introduced here. The approach is further referred to as *Simultaneous Reasoning and Mapping*, or *SRAM*. The general schematic of *SRAM* is shown in Fig. 5.2.

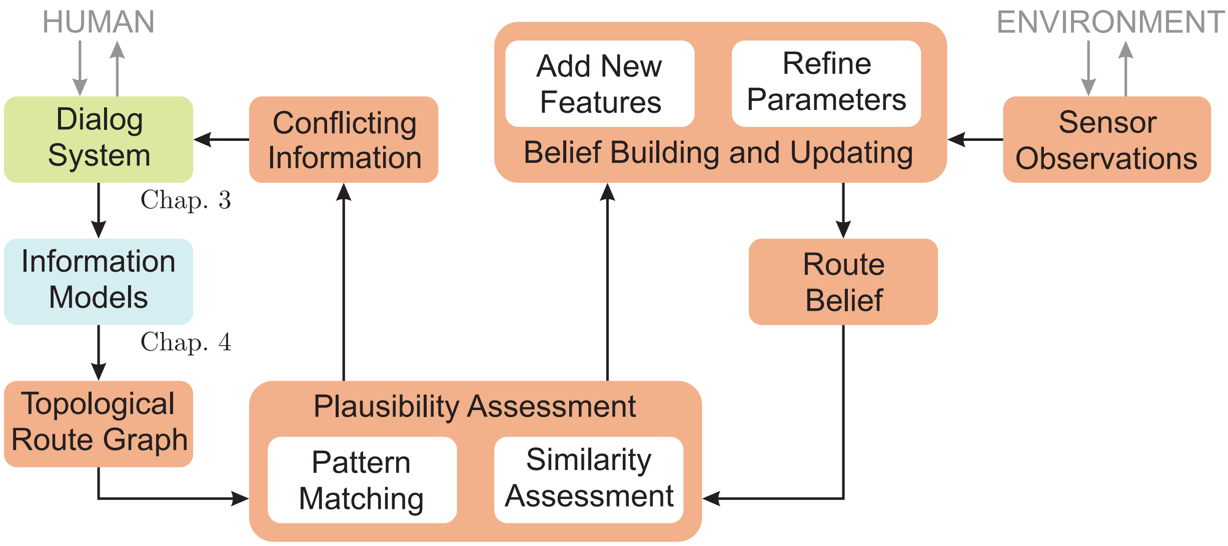


Fig. 5.2: General schematic of the *SRAM* system.

The system receives route information extracted from human-robot dialogs by a dialog system, see Chapter 3, as input. The route description is represented internally as a topological route graph applying probabilistic models for route information presented in Chapter 4. The route graph is then compared to existing route belief by pattern matching and similarity assessment in order to cover all kinds of possible errors listed above. If necessary the dialog system is caused to inquire about information that conflicts with existing information. Finally, plausible route information is included in the route belief, while implausible information is discarded. The individual modules of the *SRAM* system are described in detail below.

5.3.1 Information Representation by Topological Route Graphs

The dialog system, as in Chapter 3, asks humans for directions and extracts route descriptions from human-robot dialogs. Information extracted by the dialog system includes a sequence of route segments that are delimited by decision points and defined by directions and distances. Within route descriptions the directions represent the actions that are necessary to complete a route towards a goal. The distances between decision points refine the actions in terms of their duration. Individual actions are delimited by landmarks, i.e. decision points such as intersections, traffic lights, or buildings.

Information on individual route segments can be distorted, simplified, and erroneous. Therefore it is represented probabilistically. Probabilistic models for direction and distance information are introduced in Chapter 4. The probabilistic models allow robots to assess the accuracy and reliability of direction information by certainty values and of distance information by posterior probabilities. The certainty values are included in the topological route graphs when representing route descriptions and the posteriors can be used during navigation to assess whether a location corresponds to a description.

Extracted route information R and the route belief B of the robot are represented as directed topological route graphs. A topological route graph $G\langle N, E\rangle$ includes nodes $N_i(l_i)$ representing landmarks of type l_i along the route and edges $E_k(N_i, N_j, \delta_k, d_k, c_k)$ representing actions connecting the landmarks. Edges hold information about the direction δ_k from node N_i to N_j, i.e. the given direction information is represented as an angle relative to the previous direction, and the distance d_k between the respective nodes N_i and N_j. The landmark types l_i, l_j and the distance d_k between the landmarks, i.e. the nodes, are optional information, as they are not necessarily provided explicitly by the human. They are represented by default values, if missing. An edge E_k additionally holds a certainty value c_k reflecting the fact that the information is extracted from human route descriptions, where miscommunications or misunderstandings [109] can occur, or the provided route description may be inaccurate or even erroneous. The certainty value c_k is calculated as described in Chapter 4 in (4.13), as a weighted average of a direction probability and an error probability. It takes into account the direction of the previous route segment, the direction of the overall previous route, and the probability that an error occurs depending on the number of route segments k. The certainty value c_k is usually higher for small k, and decreases with the number of route segments. The certainty value is assigned initially when information is extracted. It is updated when new information is combined with the route belief, as discussed below.

Both new route descriptions and existing route belief are represented as topological route graphs. Therefore they share a common structure and can be compared in the plausibility assessment module. After it has been extracted and translated into a route graph, a route description is assessed for plausibility by comparing it to the existing route belief. Plausibility assessment consists of pattern matching and route similarity analysis, in order to cover all types of differences listed in Section 5.2. Both methods use a vector similarity function to compare two vectors.

5.3.2 Vector Similarity

As the main component of route descriptions are directions and distances, individual route segments and whole route graphs can be represented as vectors including both characteristics. Therefore a possibility for comparing route graphs is a suitable similarity measure for vectors. Both of the two properties magnitude and direction have to be taken into account when assessing the similarity between vectors. A magnitude similarity and a direction similarity are introduced, and then superposed to obtain a vector similarity function.

When introducing a similarity metric or function, basic properties must be satisfied. A similarity metric $S(\boldsymbol{n}, \boldsymbol{m}) : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ satisfies the properties shown in Table 5.1: non-negativity and reflexivity which guarantee positive definiteness, symmetry which ensures an unbiased comparison of two characteristics, and finally the triangle inequality. If the triangle inequality is not satisfied, but all other properties hold, the weaker measure is not called a similarity metric, but a similarity function.

Tab. 5.1: Properties of similarity metrics.

Non-negativity	$0 \leq S(\boldsymbol{n}, \boldsymbol{m}) \leq 1$
Symmetry	$S(\boldsymbol{n}, \boldsymbol{m}) = S(\boldsymbol{m}, \boldsymbol{n})$
Triangle inequality	$S(\boldsymbol{n}, \boldsymbol{m}) + S(\boldsymbol{m}, \boldsymbol{k}) \leq 1 + S(\boldsymbol{n}, \boldsymbol{k})$
Reflexivity	$S(\boldsymbol{n}, \boldsymbol{m}) = 1$ iff $\boldsymbol{n} = \boldsymbol{m}$

In the following a vector similarity function is introduced based on a direction similarity metric and a magnitude similarity function. Direction similarity of vectors $\boldsymbol{n}$, $\boldsymbol{m} \in \mathbb{R}^2$ with angles δ_n and δ_m respectively is defined as

$$S_D(\boldsymbol{n}, \boldsymbol{m}) := \left| 1 - \frac{|\delta_n - \delta_m|}{\pi} \right| , \tag{5.1}$$

with $\delta_n, \delta_m \in [0, 2\pi[$. It can be shown that all four properties for similarity metrics in Table 5.1 hold. Therefore the presented direction similarity S_D is a similarity metric. It reflects the similarity of the directions of two vectors.

Magnitude similarity of vectors $\boldsymbol{n}$ and $\boldsymbol{m}$ with magnitudes $\|\boldsymbol{n}\| = \sqrt{n_x^2 + n_y^2}$ and $\|\boldsymbol{m}\| = \sqrt{m_x^2 + m_y^2}$ is defined as

$$S_M(\boldsymbol{n}, \boldsymbol{m}) := 1 - \left| \frac{\|\boldsymbol{n}\| - \|\boldsymbol{m}\|}{\|\boldsymbol{n}\| + \|\boldsymbol{m}\|} \right| . \tag{5.2}$$

The magnitude similarity S_M satisfies the properties of non-negativity, symmetry, and reflexivity, but not the triangle inequality, therefore it is a similarity function. It describes the similarity between directions of two vectors.

A vector similarity function that accounts for both the similarity in magnitudes and the similarity in directions of two vectors is obtained by the superposition of the direction similarity in (5.1) and the magnitude similarity in (5.2) as

$$\begin{aligned} S_V(\boldsymbol{n}, \boldsymbol{m}) &:= S_D(\boldsymbol{n}, \boldsymbol{m})\, S_M(\boldsymbol{n}, \boldsymbol{m}) \\ &= 1 - \left|\frac{\delta_n - \delta_m}{\pi}\right| - \left|\frac{\|\boldsymbol{n}\| - \|\boldsymbol{m}\|}{\|\boldsymbol{n}\| + \|\boldsymbol{m}\|}\right| + \left|\frac{\delta_n - \delta_m}{\pi}\right| \left|\frac{\|\boldsymbol{n}\| - \|\boldsymbol{m}\|}{\|\boldsymbol{n}\| + \|\boldsymbol{m}\|}\right| . \end{aligned} \tag{5.3}$$

The vector similarity S_V is a similarity function, as it satisfies the properties of non-negativity, symmetry, and reflexivity, but does not satisfy the triangle inequality.

The vector similarity function can be applied to assess the similarities of both individual route segments, represented as vectors along edges, in pattern matching and whole routes, represented by a vector connecting the start and end nodes, in route similarity assessment.

5.3.3 Pattern Matching

To identify identical route graphs or gaps, excess, or rotations of individual route segments, new route graphs are compared to the existing route belief. Since the compared route graphs are simple connected graphs they cannot be matched by comparing the connectivity of the nodes as in classical graph-matching approaches, but have to be compared in terms of the properties of the edges E_k by pattern matching. Pattern matching is applied to compare the sequence of edges in the new route graph R to the sequences of edges of all paths A_i in route belief B.

Pattern or string matching is employed to find a certain pattern sequence of length m in a sequence of length n, with $n \geq m$. Typical pattern matching applications are analyzing genetic sequences [163] in computational biology and text search [192] in signal processing. The problem of identifying similar sequences or patterns, potentially with slight differences, gaps, or excess, is tackled by approximate pattern matching [4, 130]. Approximate pattern matching identifies similar patterns by calculating a matching metric and comparing it to a given threshold. A typical matching metric is the Levenshtein distance [104]. This metric measures the amount of difference between two sequences as the minimum number of substitutions, insertions, and deletions necessary to transform one string into the other.

In the *SRAM* system a matching metric similar to the Levenshtein distance is used to solve the pattern matching problem. The matching metric $D_M(R, A_i)$ used to compare the different route graphs R and A_i is based on the Levenshtein metric. However, $D_M(R, A_i)$ comprises not only the number of necessary substitutions, insertions, and deletions, but also the number of matching sub-sequences. The matching metric is defined as

$$D_M(R, A_i) = \sum_{j=1}^{m} e(M_{i,j})\,, \tag{5.4}$$

with the number of edges $e(M_{i,j})$ in sub-match $M_{i,j}$. In a first step, the minimum number of sub-matches between route R and a path A_i in the belief is identified. Then a solution to the pattern matching problem

$$D_M(R, A_i) > \tau_M \tag{5.5}$$

with a matching threshold τ_M is searched for. The pattern matching threshold is chosen heuristically, as a percentage of the number of edges in R. In the following this process is described in detail.

All edges in R are compared sequentially to all successive edges within all paths A_i of B. The paths A_i are all paths between the current node and the goal node. They can easily be found by a breadth first search in B. Each pair of edges is compared as vectors $\boldsymbol{r}_k$ and $\boldsymbol{a}_{i,j}$ by the vector similarity function in (5.3).The length of the compared vectors is δ_{r_k} and $\delta_{a_{i,j}}$ respectively, if both values are provided or unit length otherwise. A vector similarity matrix $\boldsymbol{S}_V$ is computed for each path such that

$$\boldsymbol{S}_V(R, A_i) = \begin{bmatrix} S_V(\boldsymbol{r}_1, \boldsymbol{a}_{i,1}) & \cdots & S_V(\boldsymbol{r}_n, \boldsymbol{a}_{i,1}) \\ \vdots & \ddots & \vdots \\ S_V(\boldsymbol{r}_1, \boldsymbol{a}_{i,m}) & \cdots & S_V(\boldsymbol{r}_n, \boldsymbol{a}_{i,m}) \end{bmatrix} . \tag{5.6}$$

Diagonals of $\boldsymbol{S}_V$ reveal whether R is identical to A_i; whether it is contained in A_i completely; or with gaps or excess; or some edges in R conflict with A_i; or R does not match A_i at all. Sub-matches $M_{i,j}$ are identified in the matrix as sequences of high values in the same diagonal. The minimum number of edges $e(M_{i,j})$ in sub-match $M_{i,j}$ is defined by a threshold τ_e, such that

$$e(M_{i,j}) > \tau_e . \tag{5.7}$$

High vector similarity values are defined as being above a certain threshold, with

$$S_V(\boldsymbol{r}_k, \boldsymbol{a}_{i,j}) > \tau_{Sv} . \tag{5.8}$$

The threshold τ_{Sv} ensures that not only edges with exactly the same directions are mapped, but also that more specific directions such as *'veer left'* or numeric angles can be mapped to similar corresponding directions.

The algorithm that matches the patterns starts by searching for the longest pattern match $M_{i,j}$ between graphs R and A_i, i.e. the longest sequence of similarity values satisfying (5.7) and (5.8) in any diagonal. The certainty values of the individual route segments are respected by starting the search at diagonals with higher certainty values, usually the main diagonal. After that it restricts the search space to avoid redundant pattern matches. The search space in the matrix $\boldsymbol{S}_V$ is restricted to the block matrices up left of the first value in a match $M_{i,j}$ and below right of the last value of $M_{i,j}$. The next pattern has to lie at least partially within the resulting search space. The search spaces of several matches are overlaid to create an overall search space.

The relative positions of the sub-matches, i.e. the numbers of the diagonals $\gamma(M_{i,j})$, hold information about missing, surplus, or conflicting route segments, as listed below. The sub-match following sub-match $M_{i,j}$ is denoted $M_{i,j+1}$.

- $\gamma(M_{i,j}) = \gamma(M_{i,j+1})$: Differing route segments are identified, by two subsequent sub-matches that are located on the same diagonal. The number of entries in the diagonal between the two sub-matches denotes the number of differing route segments.

- $\gamma(M_{i,j}) < \gamma(M_{i,j+1})$: Excess route segments in R are identified, if the number of the diagonal of a successive sub-match is higher than that of the previous sub match.

- $\gamma(M_{i,j}) > \gamma(M_{i,j+1})$: Route segments missing in R are identified, when the number of the diagonal of the latter sub-match is lower than that of the previous one.

The composition and interpretation of vector similarity matrices is depicted in Fig. 5.3, with three different pairs of compared routes A_i and R. The differences between the routes are highlighted and the respective similarity matrices $\boldsymbol{S}_V$ are depicted below the routes. Pattern matches $M_{i,j}$ are encircled in the matrices. The search spaces are marked in white while the restricted spaces marked in darker shades demonstrate the overlap of different search spaces.

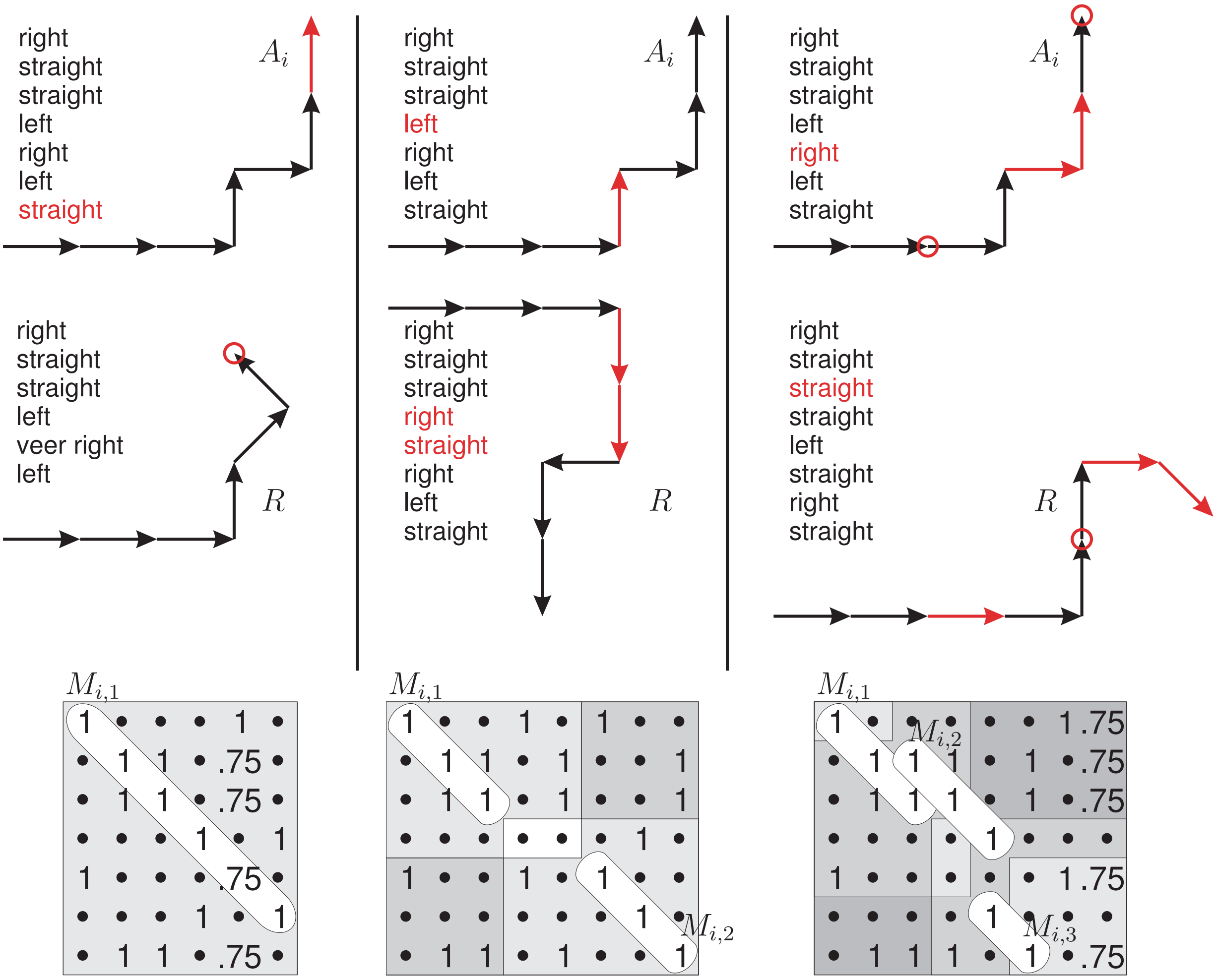


Fig. 5.3: Example graphs on the top (differences highlighted in lighter color) and respective vector similarity matrices $\boldsymbol{S}_V$ on the bottom. Sub-matches $M_{i,j}$ are encircled in $\boldsymbol{S}_V$ and search spaces and restricted spaces are marked in different shades.

A complete match $M_{i,1}$ of route R with the path A_i of belief B is shown on the left of Fig. 5.3. The center of Fig. 5.3 shows two sub-matches, $M_{i,1}$ and $M_{i,2}$. As $\gamma(M_{i,2})$ is higher by one than $\gamma(M_{i,1})$, route R holds one edge more at that position than A_i. Additionally, there is one entry missing between the matches which indicates that at that position one edge of R is different from the corresponding edge in A_i. As can be seen, there are no more matches within the search space. On the right of Fig. 5.3 three pattern matches $M_{i,1}$, $M_{i,2}$, and $M_{i,3}$ are shown. As $\gamma(M_{i,2})$ is higher than $\gamma(M_{i,1})$, again R holds one more edge than A_i. Route R has two edges less than A_i between the last two sub-matches, as $\gamma(M_{i,3})$ is lower than $\gamma(M_{i,2})$ by two. A sufficient pattern match is identified if problem (5.5) is solved, i.e. if the number of edges in all sub-matches is above a certain threshold.

The pattern match over all paths A_i in B is analyzed further. If a complete pattern match is found, the route graph R is forwarded to be included in the route belief B. Otherwise if partial pattern matches are found with gaps, excesses, or rotations, the conflicting information is sent to the dialog system and inquired about. If the conflicting information is corrected, the route graph is revised and propagated to be included in the route belief B, as well. Route information that has not been found plausible by pattern matching in this way, is assessed for vector similarity as described below. This permits the identification of other possibly plausible route information, i.e. a different route leading to the same goal.

5.3.4 Route Similarity Assessment

A vector similarity assessment between route graph R and all paths A_i in belief B is performed if no, or only an insufficient, pattern match is found, i.e. problem (5.5) is not solved. The vector $\boldsymbol{R}$ spanned by start node and end node of graph R is computed, assuming unit lengths of all edges. It is compared to the vectors $\boldsymbol{A}_i$ between the current node to all reachable end nodes, i.e. nodes with the landmark type given as *'goal'*, in B by the vector similarity function (5.3). If the highest similarity value of the compared vectors exceeds a threshold τ_S, the route graph R is assumed to be plausible. Otherwise it is assumed to be implausible and is discarded. The condition for plausibility by route similarity assessment is given as

$$\max_k \left(S_V(\boldsymbol{R}, \boldsymbol{A}_i)\right) > \tau_S \, . \tag{5.9}$$

If the new route information R is assessed to be plausible by route similarity assessment, it is assumed to be a different, previously unknown route to the same goal, and is included in the route belief B.

5.3.5 Inquiry about Conflicting Information

If a conflict between the new route graph R and the current route belief B is identified during pattern matching; the human partner is asked about this conflict by the dialog system. Conflicts C^R that are inquired about are missing, surplus, or differing route segments at the beginning or end of the route, or between consecutive sub-matches.

The conflicting information C^R, i.e. the gap in front of, between, or after sub-matches, is presented to the human along with the identified alternative information C^B from the route belief of the robot. The human partner is asked to confirm or to correct the conflicting information C^R. According to the human's answer the conflicting information is either retained or replaced by the respective information C^B from the route belief. In the case that more than one conflict is found, the next conflicting information is inquired about in the same way. After confirming or replacing all conflicts, the new route graph R is included in the belief B, possibly adding new nodes and edges.

This procedure gives the human the possibility of correcting errors in the extracted route description, but also the possibility of confirming a differing route description that either gives an alternative to or provides a specification of an existing route section.

5.3.6 Building and Updating Route Belief

The route belief of a robot represents the route information internally and serves as a basis for navigation and global path planning. It is constructed in the *SRAM* system from plausible route information and augmented by metric sensor data from the real environment.

New route information in the form of a route graph R is assessed for plausibility by pattern matching and route similarity assessment and eventually discarded or included in the route belief B of the robot. This process is depicted as a flowchart in Fig. 5.4.

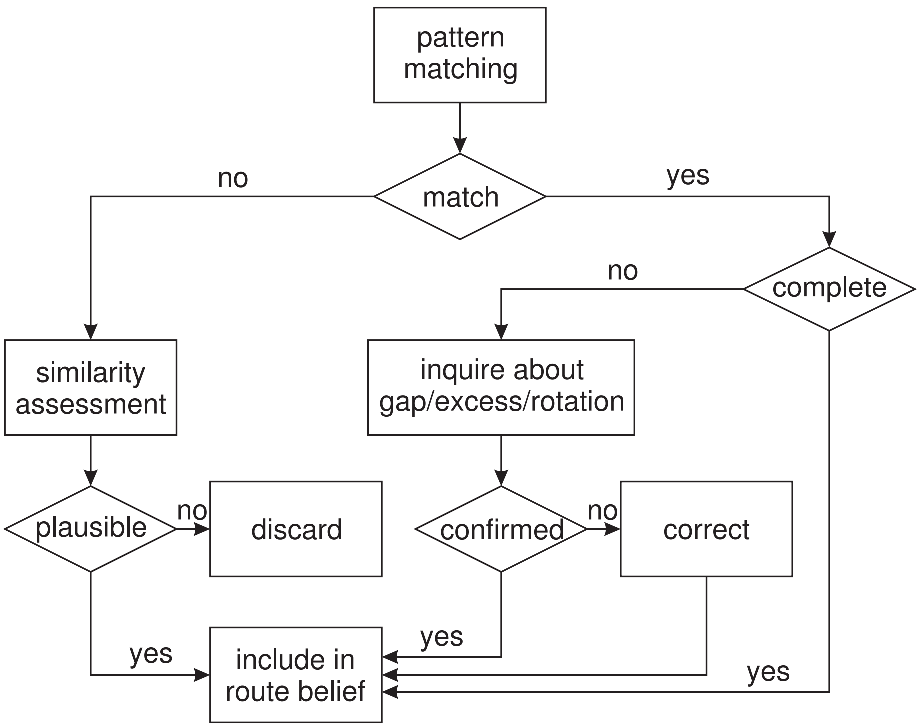


Fig. 5.4: Flowchart of the processes of plausibility assessment and route belief building.

Route information that has been found plausible by the plausibility assessment module is included in the route belief of the robot. The route to be included and the information about which edges of route R match which edges of belief B are forwarded to the route belief building and updating module.

The edges of R are included in B successively. If an edge E_k of R does not match an edge E_l in B, it is added to the belief B including the direction δ_k, the certainty value c_k, distance d_k, and the respective landmark types l_i and l_j.

If an edge E_k of R corresponds to an edge E_l in B, the properties of edge E_l are updated. Existing landmark types l_i and l_j are added to the list of landmark types of the nodes N_i and N_j corresponding to edge E_l. If one of the matching edges holds distance information this information is assigned to edge E_l. If both of the edges from R and B hold a distance information, the values are combined as a weighted average where the weights are provided by the relative posterior probabilities calculated as in (4.15) presented in Chapter 4. The certainty value c_l of edge E_l in belief B increases with

$$c_l = 1 - (1 - c_k)(1 - c_l), \tag{5.10}$$

where c_k is the certainty value of edge E_k in R. In this way the route belief is refined and supplemented with each new plausible route graph.

Initialization of Route Belief

When the system starts without any route belief, i.e. $B = \emptyset$, new route information cannot be validated and assessed for plausibility at first. Therefore initially a preliminary route belief B_{prel} is generated. As long as the belief is empty, new route information is translated into a topological route graph as described above and stored as the preliminary route belief $B_{\text{prel},i}$. When a new route description is extracted from a dialog with another human partner, the corresponding topological route graph is saved in the preliminary belief as $B_{\text{prel},i}$ and compared to the existing preliminary route graphs $B_{prel,1}$ to $B_{\text{prel},i-1}$. If two of these route graphs are found to be plausible by pattern matching or similarity assessment, they are combined as described above to form the route belief B. The remaining preliminary beliefs $B_{prel,j}$ are classified as implausible and therefore discarded. If all the graphs $B_{prel,1}$ to $B_{prel,i}$ are in conflict with each other, more information is extracted through communication with other humans and plausibility assessment is applied.

Augmenting the Route Belief with Metric Data

The route belief acquired by the *SRAM* system in the way introduced above can be used by the robot to navigate along a route towards the designated goal location. The best path A_i in the belief B to navigate is the one with the highest certainty values of the individual edges, and simultaneously the shortest path. The optimization problem is

$$\max_i \left(\frac{\sum_k c_k^{(i)}}{\sum_k d_k^{(i)}} \right), \tag{5.11}$$

where the lengths d_k equal one if not all edges in B hold a distance information. This optimization guarantees a high probability to reach the goal with low travelling costs. The most certain shortest path in (5.11) can be found for example by the Dijkstra Algorithm [37].

During navigation the robot collects sensor observations which are used to augment the route belief. Sensor observations present the system with metric information about the environment. Values in the route graph that can be augmented by sensor information are the direction and the distance, as well as recognizable landmark types. The respective edge E_k is added metric information E_k^M on top of the topological information E_k^T. The metric edges $E_k^M(N_i(l_i^M), N_j(l_j^M), \delta_k^M, d_k^M, c_k^M)$ include the gained sensor information direction δ_k^M, distance d_k^M, and landmark types l_k^M. The metric layer additionally holds a certainty value c_k^M, as the measured information is extracted probabilistically by imperfect sensors. The certainty value is set to the confidence value c_k^{sensor} of the processed sensor data

$$c_k^M = c_k^{\text{sensor}} . \tag{5.12}$$

In this way the robot validates the information extracted from human route information and adds metric data which corresponds to the layers in the SSH [91].

5.4 Evaluation

The *SRAM* system presented above was evaluated by six participants who communicated with the dialog system and gave route directions. The participants were asked to memorize the route between the robot and the goal marked on a city map, as depicted on the left side of Fig. 5.5. On the next day they each interacted with the system and gave directions which were extracted by the dialog system. The participants were all given the same start point and orientation in the map to keep the experimental conditions constant. The system works in the same way if the start point is changed while the robot navigates.

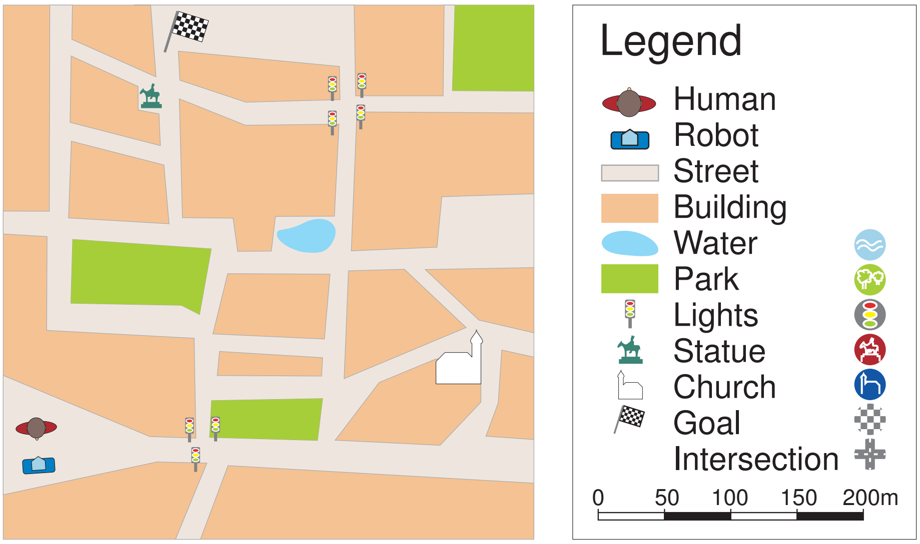
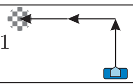
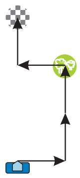
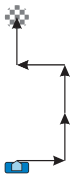
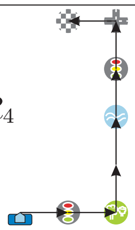
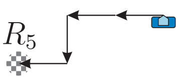
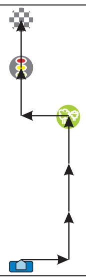


Fig. 5.5: City map used for the experimental evaluation of the *SRAM* system.

The initial route graphs R_1 to R_6 extracted from the route descriptions, the information given by the participants, and the internal processes of the *SRAM* system are presented in Table 5.2 for each new participant. Initially, both route graphs built from the route descriptions R_1 and R_2 are not found to be plausible and are therefore stored in a preliminary belief. Route description R_1 is correctly identified as implausible and is discarded when the third participant gives a route description and route information R_2 and R_3 are combined to form the belief B. Route R_4 is different from the other routes, but is classified as plausible by the route similarity assessment module, as the maximum vector similarity $S_V(\boldsymbol{r}, \boldsymbol{b}) = 0.91$ between the vectors from the start nodes to the end nodes of the new route R and the belief B is higher than the chosen threshold. On the other hand route description R_5 is identified as plausible but missing the first route segment. It is corrected after inquiring about the missing route segment and the human confirms that route segment of direction *'right'* needs to be inserted at the beginning of the route. Finally, route R_6 presents a refined description compared to the other route descriptions already included in the route belief. After inquiring about the differing route segments, they were confirmed by the human participant and included in the route belief B by the system. (The values of the thresholds in plausibility assessment were assigned as $\tau_M = 0.6$, $\tau_{S_V} = 0.75$, and $\tau_S = 0.8$ in the experiment.)

Tab. 5.2: Process of building a route belief B during the experiment.

Route	Human information	*SRAM* system action
R_1 	straight, left, straight (goal)	store R_1 as preliminary belief $B_{\text{prel},1}$
R_1 	right, left, straight (park), left, right, straight (goal)	compare R_2 to $B_{\text{prel},1}$ $\longrightarrow$ no match ($e(M_j) = 2$, $\max(S_V) = 0.46$)
		store R_2 as preliminary belief $B_{\text{prel},2}$
R_3 	right, left, straight, left, right, straight (goal)	compare R_3 to $B_{\text{prel},1}$ $\longrightarrow$ no match ($e(M_j) = 2$, route sim. $= 0.46$)
		compare R_3 to $B_{\text{prel},2}$ $\longrightarrow$ match ($e(M_j) = 6$)
		store R_3 with $B_{\text{prel},2}$ as belief B
		discard implausible $B_{\text{prel},1}$
R_4 	right (lights), straight (park), left, straight (water), straight (lights), straight (intersection), left (goal)	compare R_4 to B $\longrightarrow$ plausible ($\max(S_V) = 0.91$)
		combine R_4 with B
R_5 	left, straight, left, right, straight (goal)	compare R_5 to B $\longrightarrow$ plausible ($e(M_j) = 5$)
		'Should direction 'right' be inserted at the beginning?'
	'yes'	correct route graph R_5
		combine R_5 with B
R_6 	right (lights), left, straight, straight (park), left, right, straight (lights), straight (goal)	compare R_6 to B $\longrightarrow$ plausible ($e(M_j) = 6$)
		'Third route segment superfluous?'
	'no'	*'Last route segment superfluous?'*
	'no'	combine R_6 with B

The route belief B formed during the experiment is depicted in Fig. 5.6 in robot coordinates, with the different paths A_1 to A_3 marked in different line styles and colors. The extracted landmark types l_i are depicted by icons. A default value was inserted where no explicit landmark type was extracted.

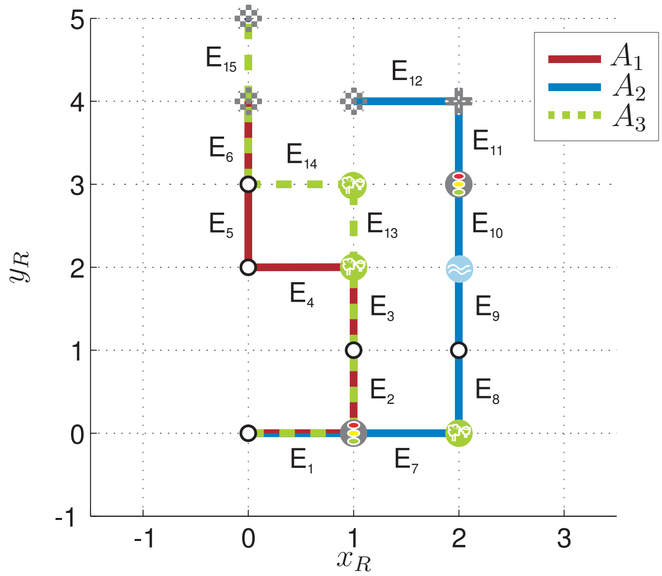


Fig. 5.6: The resulting route belief B in robot coordinates.

The certainty values of the individual edges in the route belief B over the time are depicted in Fig. 5.7. The dashed lines represent values of edges in the preliminary belief $B_{\text{prel},1}$ (denoted by points in light color) and $B_{\text{prel},2}$ (denoted by circles in dark color) respectively. Solid lines represent certainty values of edges in the route belief B. The certainty values of the edges in the preliminary belief $B_{\text{prel},1}$ drop to zero when $B_{\text{prel},1}$ is assessed as not plausible and rejected. The certainty values c_k of the edges of the belief given by four of the participants increase towards value one. While less frequently stated routes have lower certainty values. The certainty values c_{13} and c_{14} of the route segments that have been added from route description R_6 are relatively low compared to certainty values c_4 and c_5 of the alternative route segments. As route description R_6 has a higher granularity than the alternative descriptions and the human has confirmed the given route, the corresponding route segments could be assigned higher certainty values by the route building and maintenance module. Furthermore, in a scenario where a robot navigates along the route towards the goal and asks for directions along the way, the certainty values that are initially assigned to the individual edges of the route descriptions are higher for route segments close to the current position. In such a scenario the certainty values c_{13} and c_{14} of the latest route description would be higher.

The outcome of the route belief of the robot would be basically the same if the sequence of given route descriptions were interchanged. In the worst case the route descriptions R_1 and R_4 would be the first to form the preliminary belief and the faulty route R_1 would be found to be plausible as it is a complete sub-match of R_5. In this case, as a route segment is missing at the beginning of R_1, the *SRAM* system would feedback to the dialog system to clarify the conflict with the human, i.e. inquire whether a first segment of direction *'right'* has to be inserted. Either the route graph would be corrected, if the human confirmed

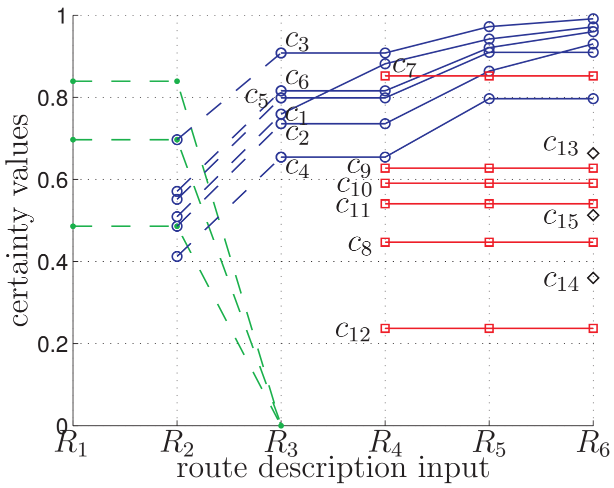


Fig. 5.7: Certainty values of all edges in belief B over the time. Dashed lines denote certainty values of edges in the preliminary belief B_{prel}.

that a segment has to be inserted, or if the human insisted on the erroneous information, the graph would be retained in the given form and added to the route belief. In that last case, the belief would contain an implausible route. However, as this route is not confirmed by other persons the certainty values would not change, while the certainty values in the other routes would increase which would therefore be selected for navigation due to the optimization of (5.11). This shows that the order in which the route descriptions are given does not make a difference to the outcome of the process in the *SRAM* system.

5.5 Discussion

Robots, navigating based on route descriptions extracted from human-robot dialogs, need to reason about these route descriptions, as they are not necessarily correct and can have different degrees of refinement. Route descriptions can include minor differences that emerge from different granularities of information, small errors that can be corrected by inquiring about them, or they can be erroneous as a whole. Therefore robots must assess the plausibility of route descriptions in comparison with other information.

This chapter presented a system for *Simultaneous Reasoning and Mapping* (*SRAM*). It provides a general procedure for reasoning about whole route descriptions extracted from human-robot dialogs, representing route information, reasoning about the information, clarifying conflicting information, and building a route belief. The system uses the dialog system presented in Chapter 3 to extract missing route information, and giving feedback to humans. The route description is represented internally using the probabilistic models presented in Chapter 4. The *SRAM* system interfaces with the systems and methods presented in the previous chapters and enables a robot to represent route information extracted from HRI as route graphs and to compare the information to the current route belief by pattern matching and vector similarity assessment, in this way identifying plausible or conflicting information. If a conflict of the extracted information with the route belief is found, the dialog system inquires about this conflicting information. A route belief is formed from all plausible and verified route information. Additionally the system adds metric information to the route belief from sensor observations during navigation.

The *SRAM* system has been tested by non-expert users in an experiment. In summary, all correct route descriptions were identified as plausible and were subsequently included in the route belief. The system correctly identified an incorrect route description as implausible and discarded it. An incomplete route was identified as such and after feedback to the human participant the route was corrected and included in the belief of the robot. A route describing a different way to the same goal was classified as plausible as well by means of vector similarity assessment and was subsequently included in the route belief. The certainty values of route segments that were given in route descriptions by several participants increase, reflecting a higher confidence in that certain information. The evaluation showed the procedure and effectiveness of the presented *SRAM* approach. The system is robust to a permuted sequence.

Other researchers have studied reasoning about spatial concepts before, however they have not tackled the problem of representing and comparing overall route descriptions and assessing them for plausibility. The presented SRAM system provides a novel framework for extracting route information, representing route information probabilistically, and reasoning about it. The major contribution of this chapter is the plausibility assessment method which consists of a pattern matching metric applied to compare different route graphs to identify similar routes possibly with slight differences and a complementary route similarity assessment method for identifying different routes with the same goal. SRAM is a novel framework that can be applied not only to extracting and reasoning about route information but other information that can be represented as graphs, e.g. relations between persons, objects, or concepts.

6 Conclusion and Outlook

This thesis investigated methods that allow robots to fill gaps in their route knowledge by asking humans for directions, reason about the extracted route information, and represent it in a route belief. In this chapter, the presented approaches are summarized and possible directions for future work are proposed.

6.1 Concluding Remarks

An essential ability for robots assisting humans in intelligent and versatile ways is to extract missing task knowledge to be able to operate in new and unforeseen situations as well as unstructured, complex, and dynamic environments. A legitimate approach for robots to come by missing knowledge in general is to ask humans for information. In the special case where a robot is given the task to navigate to a given goal within an unknown or changing environment, it can extract missing route knowledge by asking humans the way. Challenges of this strategy are proactive extraction of unambiguous information from HRI, assessing information for reliability and plausibility, as well as representing it internally in the route belief of a robot.

In this thesis methods for extracting probabilistic representations of route information from human-robot dialogs were presented. In the following the individual chapters of the thesis are summarized and the main contributions are presented.

To derive specific research questions, an experiment conducted with the interactive outdoor robot *ACE* was analyzed in Chapter 2. In the experiment, *ACE* was given the task of navigating to a designated goal location in an unknown environment, without any previous map knowledge or GPS, but solely by asking passers-by for directions. The experiment was successful in general. A few limitations are pointed out by the experimental results. These suggest further research questions, namely how to proactively extract route information from natural-language dialogs, how to model route information probabilistically to evaluate the accuracy and reliability of it, and how to assess extracted route descriptions for plausibility by comparing them to descriptions from other persons. These questions are tackled in Chapters 3 to 5. *Experiments by other researchers either dealt with autonomous robot navigation in outdoor environments or with HRI in structured indoor environments. The ACE robot was the first to interact with non-expert human users in an urban outdoor environment and extract missing route information for navigation from HRI.*

Chapter 3 focuses on natural-language direction-inquiry dialogs between humans and robots. It introduces a dialog system enabling robots to close gaps in their route knowledge by asking humans for missing information. Mechanisms from human-human communication from linguistic principles are adapted to derive guidelines for human-robot dialogs. The guidelines are implemented in the dialog system to render the human-robot communication natural and intuitive to non-expert human users and to extract unambiguous route information. The dialog system is evaluated experimentally under laboratory conditions. Subjective results from the evaluation show that the application of mechanisms from human-human communication in the form of guidelines for HRI were favorably evaluated by human users. The users perceived the dialogs as natural. At the same time objective results show that the system performance is comparable to the work of other researchers. *State-of-the-art dialog systems usually tackled communication only from a technical point of view and focused on objective measures in their evaluations where they used hand pre-selected keyword input to their systems. This system analyzes raw text input and specifically applies guidelines for HRI derived from linguistics to render dialogs more natural to humans, and close knowledge gaps by asking for missing information.*

Probabilistic models for individual direction and distance information in route descriptions are presented in Chapter 4. Information given in route descriptions such as directions and distances between decision points are approximations and simplifications of relations in the real world. In order to provide robots with mechanisms to assess the accuracy and reliability of it, such information is modeled probabilistically. Models for direction information consider the direction of the previous route and the possibility of errors in the route description, resulting in certainty values that reflect the reliability of the direction information in each segment of a route description. Distance information in route descriptions is modeled as posterior probabilities of real distances, given the estimated distance, in order to provide robots with the means of assessing the accuracy of such information. The validity of the models is demonstrated by representing an example route description probabilistically including certainty values and probability distributions of directions and distances of all route segments. The evaluation confirms that the presented models provide reasonable assessment of individual route information. *The presented probabilistic models of route information provide a means to assess the accuracy and reliability of given information. While models of cognitive directions and distances have been found by other researchers, models for information given during route descriptions which differ from the former have not been derived before. The presented probabilistic route information models are not only novel in the field of spatial cognition, but find a meaningful application in robotics as well.*

Chapter 5 presents an approach to assess entire route descriptions for plausibility. Just like direction or distance information of individual route segments, whole route descriptions are prone to simplifications, distortions, and errors. Thus, different routes from descriptions by different persons are compared in order to identify plausible and implausible routes and possibly correct conflicting information. A method for *Simultaneous Reasoning and Mapping* (*SRAM*) is presented. The developed *SRAM* method assesses routes for plausibility by pattern matching and route similarity assessment, gives feedback to the human to inquire about conflicting information if necessary, and builds a graph-based route belief from plausible routes. The *SRAM* system links the dialog system in Chapter 3 and the probabilistic models in Chapter 4 in a framework for reasoning about route information and representing it. The effectiveness of the *SRAM* approach is shown in an experimental evaluation where the system identified all correct and erroneous route descriptions correctly and built a plausible route belief. Additionally, the system exhibits robustness to interchanging the sequence of route description input. *The SRAM system provides a sound method for reasoning about information from different sources before including it into a route belief. No comparable approach has been presented by other researchers before. This system is necessary for robots to have the means of comparing different information, inquiring about conflicting information, and including sound information into the belief.*

In a nutshell, this thesis provides robotic systems with methodologies for explicitly asking humans for missing route information and building a route belief as a probabilistic internal representation of the extracted information. *The presented approaches go beyond the state of the art: they allow robots not only to proactively extract information from human-robot dialogs by explicitly asking for missing knowledge, but also to reason about the extracted information; means of probabilistically representing individual route information in route descriptions are provided; finally, SRAM presents a novel system for simultaneously reasoning about different route descriptions and building a route belief as an internal representation.* The presented methods are expandable and generalizable to extracting and reasoning about other types of missing information.

6.2 Outlook

The research presented in this thesis is a first step towards general information extraction and belief construction through human-robot communication. In light of the broad field of involved research topics and applications there are still a number of open research questions and interesting future directions.

This thesis focuses on the specific problem of extracting a probabilistic representation of route information from descriptions given by humans. In general, robots operating in unstructured, complex, and dynamic human-populated environments need a wide range of information to be able to complete their tasks. Therefore it would be useful to enable them to extract and represent any kind of missing task information. For this purpose the presented methods have to be extended and adapted to the respective applications.

Additionally, the presented methods have applications in the currently growing research areas of service robotics and multi-robot systems. In these fields interaction with non-expert human users and between humans and multi-robot systems are upcoming challenges which will bring along new research questions.

- The developed dialog system is expandable to ask for any kind of missing task knowledge. Many of the derived guidelines for human-robot dialogs are applicable for general discourse topics, while more guidelines have to be deduced from linguistics for different dialog contents. These guidelines have to be integrated in robot dialog systems in order to overcome the vagueness of natural language and render the dialogs natural to humans.

  In order to render dialogs with robots even more natural and intuitive to humans, other means can be applied. Robots would profit from the ability to identify, mend, and ideally avoid possible miscommunications in human-robot dialogs inspired by human-human communication. Other strategies that may further improve the communication between humans and robots include rhetoric, robot body language, proxemics, and emotional behavior.

In multi-robot scenarios, thought has to be given to how information is extracted and to which agents to allocate which tasks. Several robots could extract information from different humans simultaneously. Or different agents might be allocated tasks of asking different humans for different information. Such an approach would be time efficient and would provide the multi-robot system with a lot of information which can be stored and processed centrally.

- The presented probabilistic models are valid for route information in urban environments with an appropriate distance range. These models could be extended to include the effects of contextual and personal factors, and thereby become applicable to a wide range of contexts and adaptable to a variety of environmental scopes interaction partners.

  To give credit to the simplified and distorted character of any kind of information extracted from human-robot dialogs, probabilistic models need to be found not only for route information, but for other types of information as well.

- The presented *SRAM* architecture is generalizable and can be applied to other information, as long as it can be represented in graph structures. It can be used to compare and assess any kind of complex information coming from different sources.

  The *SRAM* system is designed to compare information from different humans. In this way it could be used in multi-robot systems to compare all information extracted by different robots. In such a scenario, *SRAM* can be applied to build a belief consisting of plausible information that would be usable by the whole multi-robot system.

These possible extensions and applications provide many directions in different research fields. Each of these directions requires an even closer collaboration between engineers, computer scientists, psychologists, linguists and other researchers.

A The Autonomous City Explorer

The hardware and the software of the *ACE* robot are presented here. In particular, the components and modules for the vision, interaction, and navigation systems are described.

A.1 Hardware

The *Autonomous City Explorer* robot was developed at the Institute of Automatic Control Engineering as a cooperative project of several researchers[1], as presented in [214, 224].

The *ACE* robot comprises a differential drive mobile platform with wheel encoders, developed by BlueBotics SA, two laser range finders for navigation and traversability assessment, respectively, a speaker, a touch screen, an animated mouth, as well as a sophisticated stereo vision system based on a multi-focal active camera head for image processing. The complete system measures 78 cm in length, 56 cm in width, and 178 cm in height, including the camera head, and weighs approximately 160 kg. Fig. A.1 shows the development of the robot hardware over the time, with the final design used in the experiment on the right.

Fig. A.1: Development of *ACE* over the time, from basic platform to complete system.

The mobile platform has a maximum payload of 150 kg and is moved by two wheelchair drive wheels (30 cm diameter) with differential drive and treads. It has two castor wheels (12 cm diameter) in the rear and two castor wheels on springs in the front (10.5 cm diameter). The maximum velocity is 1.4 $\frac{m}{s}$, the maximum acceleration is 1.35 $\frac{m}{s^2}$. It has an autonomy of up to 10 km depending on the paving. The climbing ability of the platform has been thoroughly tested, since this is an essential factor for outdoor navigation. The robot is capable of climbing a slope of 6° and steps of 35 mm. For urban environments

[1]The team in alphabetical order: Andrea Bauer, Martin Buss, Klaas Klasing, Kolja Kühnlenz, Georgios Lidoris, Quirin Mühlbauer, Florian Rohrmüller, Stefan Sosnowski, Dirk Wollherr, Tingting Xu.

this means that the robot can safely navigate on sidewalks and smooth surfaces but must avoid larger steps, such as the curbside.

The vision system consists of a wide-angle stereo camera mounted on a central pan/tilt-platform. The main camera used for human tracking is a 3-sensor, multi-baseline Bumblebee XB3 by Point Grey Research, with a focal length of 3.8 mm and enhanced flexibility and accuracy due the switchable baseline. A second Bumblebee stereo camera, with a focal length of 6 mm, for traffic sign detection and posture recognition is mounted underneath. The central platform is driven by DC drives with harmonic drive gears. An embedded RISC processor (MPC555, Motorola) controls the camera motions on joint levels. The camera system is encapsulated and accepts camera pose commands from a higher-level decision and planning unit via a CAN-based interface.

A.2 Software

The software is run on two onboard Linux PCs (one for navigation and interaction and one for vision processing) with four 2.2 GHz cores each, powered by an array of rechargeable lithium polymer batteries that provide power for up to 8 hours. A third PowerPC controls the differential wheel platform and receives asynchronous driving commands from the navigation PC. Processes run at fixed update rates in a pull architecture fashion, meaning data is queried from sensors and processes are refined at fixed intervals.

The system architecture is broken down into a sensor layer, a perception layer, a control layer, and an actuator layer. The three main subsystems of the *ACE* robot, for navigation, vision, and interaction, are interconnected through the different layers and each of them consists of several modules. Fig. A.2 shows the individual modules, subsystems, and layers of the *ACE* architecture.

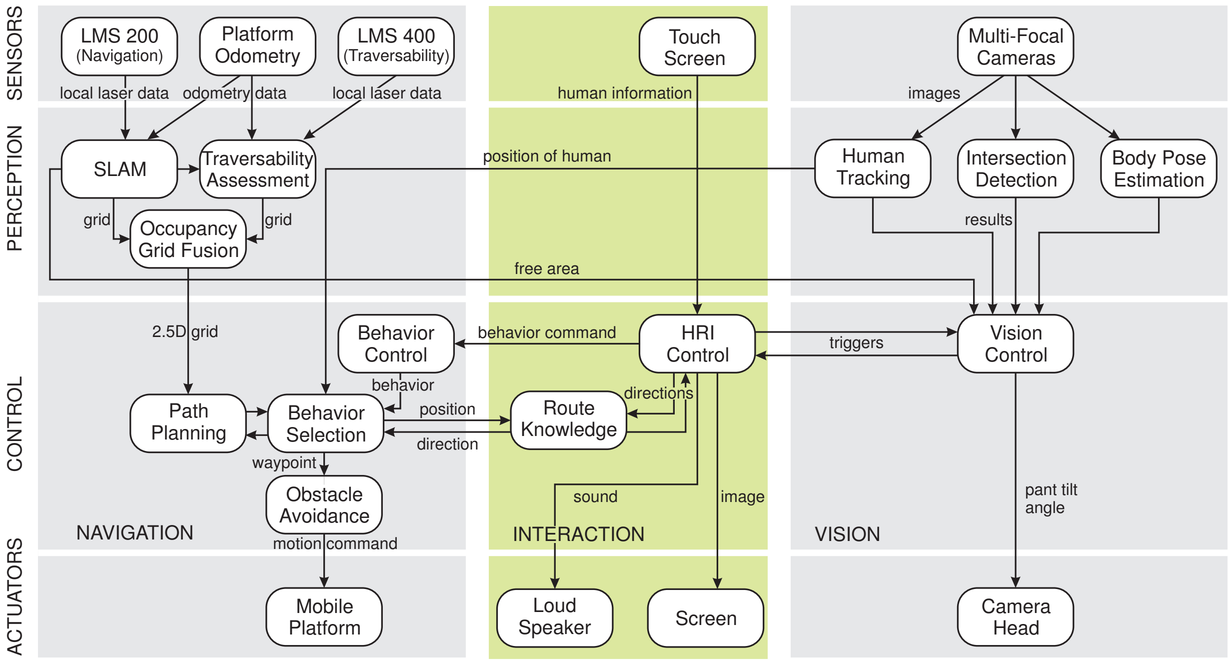


Fig. A.2: The software architecture of *ACE*, with modules, subsystems, and layers.

B Contextual Dependence of Distance Information

Distance information in route descriptions given by humans may depend on personal variables, such as gender or age, and on contextual factors, such as the scale of the environment. The contextual dependence is analyzed for quantitative and qualitative distance information.

B.1 Quantitative Distance Information

The effects of the variables gender, age, and self-assessment on the estimation of quantitative distances are assessed by statistically comparing the relative constant and absolute errors of all data points. The relative constant error of the distance estimation e_d^C computes as the relative difference between real distance and estimated distance, as

$$e_d^C = \frac{d_{\text{real}} - d_{\text{est}}}{d_{\text{real}}} , \tag{B.1}$$

while the relative absolute distance error e_d^A computes as

$$e_d^A = \frac{|d_{\text{real}} - d_{\text{est}}|}{d_{\text{real}}} . \tag{B.2}$$

The relative constant error for the time estimates e_t^C is accordingly the relative difference between real distance and estimated time multiplied by a mean walking velocity, such that

$$e_t^C = \frac{d_{\text{real}} - t_{\text{est}}\, v_{\text{walk}}}{d_{\text{real}}} , \tag{B.3}$$

with the assumed walking velocity $v_{\text{walk}} = 4.5 \frac{\text{km}}{\text{h}}$. Accordingly the relative absolute error for the time estimates e_t^A is

$$e_t^A = \frac{|d_{\text{real}} - t_{\text{est}}\, v_{\text{walk}}|}{d_{\text{real}}} . \tag{B.4}$$

The relative estimation errors for distances and times are depicted in Fig. B.1 as box-whisker plots with constant errors on the left and absolute errors on the right.

The relative constant time error e_t^C is significantly lower than the relative constant distance error e_d^C (t-test, $p = 0.0018$), as is the relative absolute time error e_t^A compared to the relative absolute distance error e_d^A (t-test, $p = 0.0017$). Thus walking time information is generally more accurate than metric distance information in a route description.

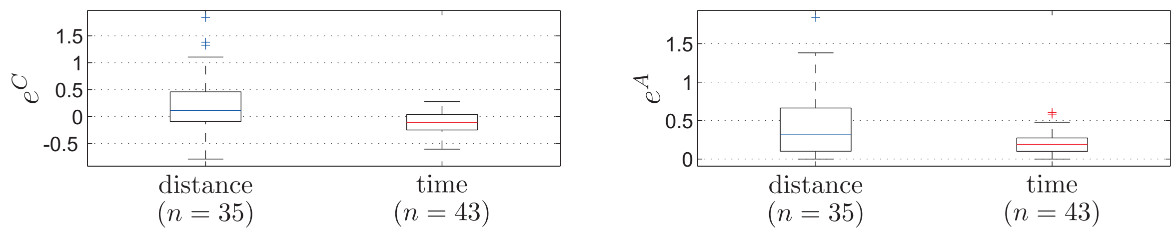


Fig. B.1: Box-Whisker plots comparing the relative constant errors on the left, and the relative absolute errors on the right, of distance and time estimates.

Effects of Gender

The survey data reveals that there are differences between genders in the preference of giving distance and time estimates. While 73.0 % of the male subjects gave a distance estimation, 60.1 % of the female subjects gave a time estimation.

A box-whisker plot for the estimation errors for male and female subjects is presented in Fig. B.2, with the errors for distance estimation on the left, and the errors for time estimation on the right.

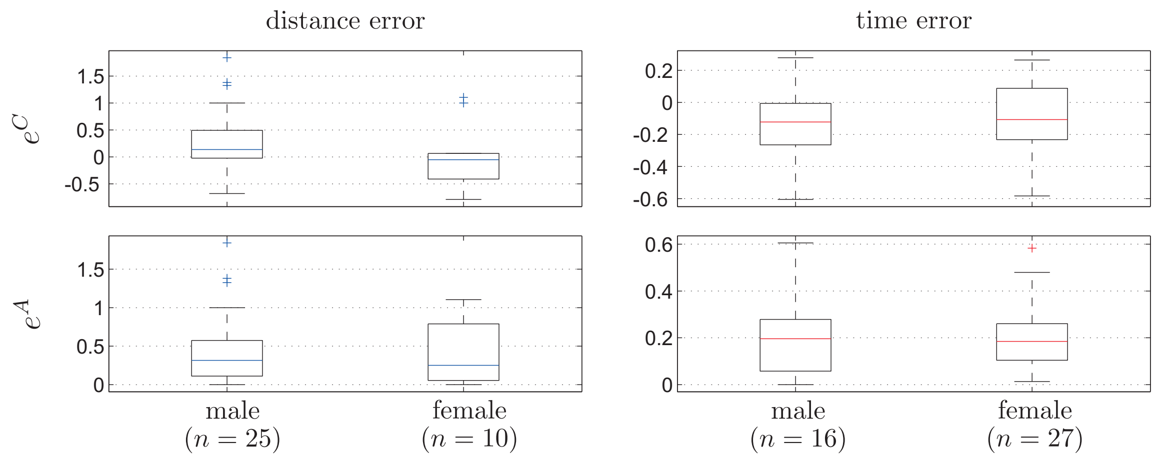


Fig. B.2: Box-Whisker plots comparing the variable gender as the relative constant errors on the top and relative absolute errors on the bottom, for distance estimates on the left and time estimates on the right.

T-tests show that the differences in the estimation errors between males and females are not significant (e_d^C: $p = 0.2078$, e_t^C: $p = 0.6911$, e_d^A: $p = 0.9122$, e_t^A: $p = 0.8501$).

Effects of Age

The subjects are divided into three age groups. To obtain adequate sample sizes, the subjects are classified into three groups, namely below 30, 30 to 49, and 50 and over. A box-whisker plot for the relative errors of distance and time estimates for the different age groups is depicted in Fig. B.3. The upper row shows the relative constant errors, while below the relative absolute errors are depicted for distance estimates on the left and time estimates on the right, respectively.

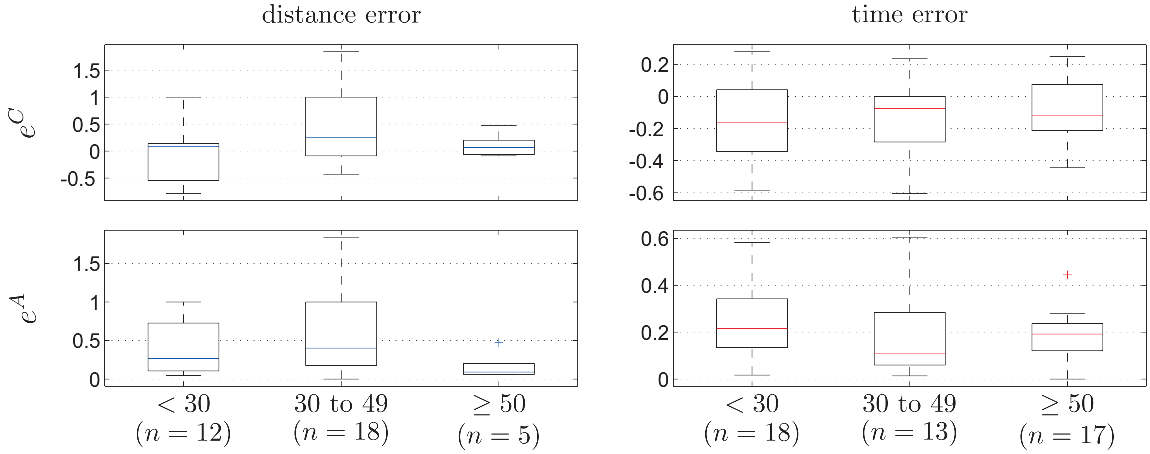


Fig. B.3: Box-Whisker plots depicting estimation errors for the variable age as the relative constant errors on the top and relative absolute errors on the bottom, for distance estimates on the left and time estimates on the right.

The estimation errors for the three age groups have been compared by ANOVAs. The differences between the groups were not significant, neither for distance estimates (e_d^C: $p = 0.0941$, e_d^C: $p = 0.1777$), nor for time estimates (e_d^A: $p = 0.8234$, e_d^A: $p = 0.6766$).

Effects of Self-Assessment

The subjects were asked to give an assessment of the accuracy of their given distance or time estimate. The answers were categorized in three self-assessment groups, namely *'bad'*, *'average'*, and *'good'*. Box-whisker plots for the estimation errors of distance and time for the variable self-assessment is depicted in Fig. B.4.

An ANOVA shows a significant difference for the relative constant distance estimation errors between the self-assessment groups (e_d^C: $p = 0.0176$). This comes about because subjects who assessed their own estimation abilities as *'bad'* tended to underestimate distances, while others tended to overestimate them. However, there are no significant differences between the relative constant time errors (e_t^C: $p = 0.1043$) or the relative absolute errors of distances and times (e_d^A: $p = 0.1742$, e_t^A: $p = 0.4129$).

The fact that differences of the relative errors are not significant for the variables gender, age, and self-assessment, except for e_d^C, shows that those personal variables have no important influence on the accuracy of walking distance or time estimation within route descriptions. Therefore these factors do not have to be taken into account when searching for and selecting a person whom to ask for directions, in order to get good distance estimations. Furthermore these personal factors are of no importance when modeling the relation between real distance and estimated distance or time.

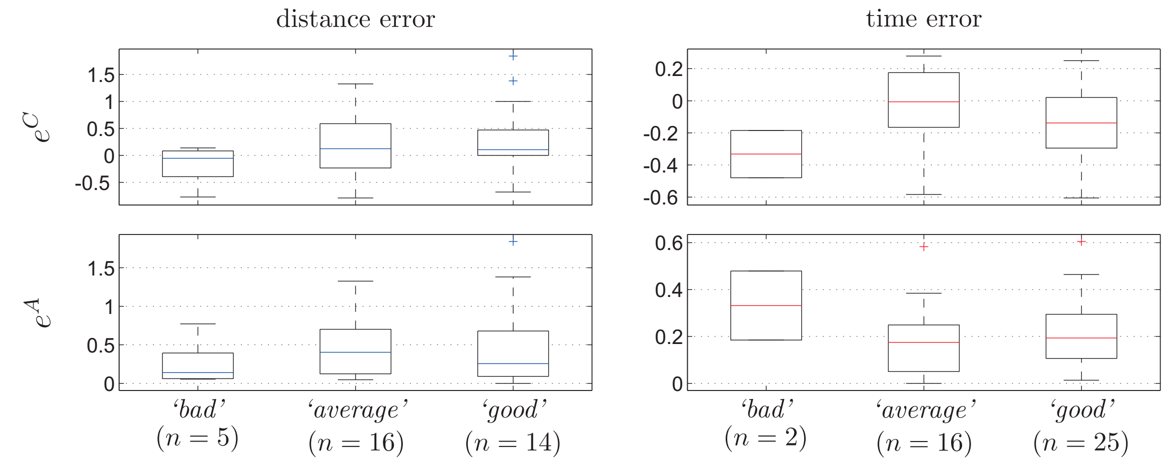


Fig. B.4: Box-Whisker plots showing estimation errors for the variable self-assessment as the relative constant errors on the top and relative absolute errors on the bottom for distance estimates on the left and time estimates on the right.

B.2 Qualitative Direction Information

An experiment was conducted to discover whether or not the deictics *‘here’*, *‘close’*, and *‘far’* describe different distance ranges in different environments. The borders between pairs of adjacent deictics, i.e. between *‘here’* and *‘close’*, and between *‘close’* and *‘far’*, are analyzed depending on the given environment, i.e. abstract, urban, or large scale environment. The participants of the experiment were 36 PhD students of engineering as participants. There were 4 female and 32 male participants aged between 24 years and 37 years, with an average age of 27 years. Each participant had to fill in questionnaires, as shown Appendix C.2, for an abstract environment, as in Fig. C.2, for an urban environment, as in Fig. C.3, and for a large scale environment, as in Fig. C.4.

The distance between start and goal point marked in the city map corresponds to an actual distance of 15 m, while the distance in the road map corresponds to a distance of 303 km in the real environment.

The test was designed with repeated measures, i.e. the sequence in which the maps were presented was alternated for every person. There were six different sequences for the three maps and every sequence of tasks was given to six participants. The participants were asked to imagine themselves traveling the depicted distance by according modes of transportation, i.e. walk in the city, drive by car between cities, and navigate the abstract distance. Subsequently they were asked to mark the borders b_c^h between *‘here’* and *‘close’* and b_f^c between *‘close’* and *‘far’* in each map on the line connecting start and goal.

Results

During the survey the two borders b_c^h and b_f^c were evaluated for each map and all participants. Participants reported that it influenced their decisions about placing the borders, when they knew the particular route, or if places were located within walking distance. They were also influenced by the fact that the depicted line did not correspond to the route they would normally take.

To compare the results from the different maps the data was scaled to a relative distance. A relative distance of zero marks the start point and 100 marks the goal point. The variances of the data differ strongly between the maps. To adjust the variances for comparability the logarithm of the data is used for further analysis. The box-whisker plots of the logarithmic data are shown in Fig. B.5.

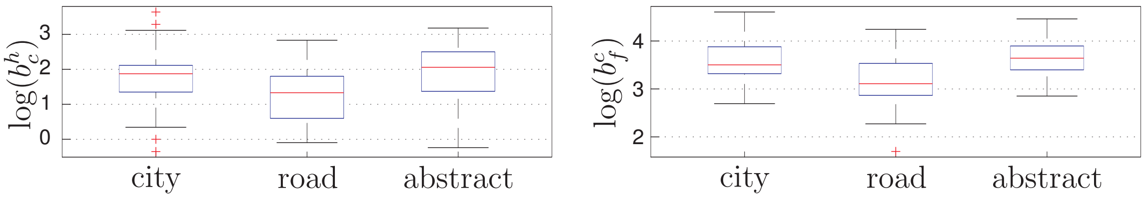


Fig. B.5: Box-whisker plots for the logarithmic borders b_c^h and b_f^c depending on the map.

A mixed linear model is fitted to the logarithmic data $\boldsymbol{y}$ to model the effects of the different maps, such that

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\epsilon} \ , \tag{B.5}$$

with the design matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$. The effects $\boldsymbol{\beta}$ of map and age, the random effects $\boldsymbol{u}$ for the subjects, and the random errors $\boldsymbol{\epsilon}$, are estimated from the data $\boldsymbol{y}$. The borders b_c^h and b_f^c are independent of each other and therefore two mixed linear models are fitted for the data of the respective borders. The random effects and the random errors are normally distributed.

The resulting fixed effects of *map* and *age* with a total of 70 degrees of freedom, estimated for the mixed linear models, are presented in Table B.1. The intercept value represents the city map.

Tab. B.1: Parameter estimates of the fixed effects in the fitted mixed linear models.

	effect	estimate	std error	p-value
log(b_c^h)	intercept	5.00	1.05	0.0000
	age	-0.11	0.04	0.0061
	road map	-0.43	0.11	0.0003
	abstract map	0.08	0.11	0.4928
log(b_f^c)	intercept	4.55	0.61	0.0000
	age	-0.03	0.02	0.1357
	road map	-0.50	0.09	0.0000
	abstract map	0.01	0.09	0.9347

The variable *road map* was statistically significant with a strong effect in both models. The variable *age* was significant in the model for b_c^h, but has only a weak effect.

The results of the survey show that the environment in which the deictics *'here'*, *'close'*, and *'far'* are used to describe a qualitative distance cannot be neglected, but must be considered when interpreting them. Also other contextual or personal factors such as the age of the person may have an effect on the meaning of these deictics. Thus for each environment different probabilistic models describe the relations between the spatial deictics *'here'*, *'close'*, and *'far'* used in a route description and the real distance.

The presented statistical analyses of conceptual dependence of distance information show that personal variables have no strong effect and do not have to be taken into account when modeling posteriors. At the same time the effect of the scale of the environment with suitable modes of transport has a strong effect on distance information in route descriptions. Thus this effect must be taken into account in distance information models.

C Questionnaires

Questionnaires were used for the subjective evaluation of the dialog system as well as for collecting data for modeling qualitative distance information in route descriptions.

C.1 Questionnaire for Dialog-System Assessment

The questionnaire in Fig. C.1 was used to gain a subjective evaluation of the dialog system presented in Section 3.5. Here the original German version of the questionnaire is presented which was used for subjective system assessment.

Im Folgenden sehen Sie einige Fragen zum Dialogsystem.
Bitte Kreuzen Sie die entsprechende Antwort an.

	Stimme voll zu				Stimme nicht zu
Das System war einfach zu benutzen.	(1)	(2)	(3)	(4)	(5)
Das System hat verstanden, was ich eingegeben habe.	(1)	(2)	(3)	(4)	(5)
Die Dauer der Interaktion war angemessen.	(1)	(2)	(3)	(4)	(5)
Ich wusste immer, was ich dem System eingeben konnte.	(1)	(2)	(3)	(4)	(5)
Das System hat reagiert, wie ich es erwartet habe.	(1)	(2)	(3)	(4)	(5)
Das System hat meine Wegbeschreibung auf Anhieb richtig verstanden.	(1)	(2)	(3)	(4)	(5)
Wenn nicht: Das System hat meine Wegbeschreibung nach der Korrektur richtig verstanden.	(1)	(2)	(3)	(4)	(5)
Die Struktur des Dialogs war sinnvoll.	(1)	(2)	(3)	(4)	(5)
Es war mir klar, aus wessen Perspektive ich den Weg beschreiben sollte.	(1)	(2)	(3)	(4)	(5)
Es ist mir wichtig, dass die Perspektive der Wegbeschreibung geklärt ist.	(1)	(2)	(3)	(4)	(5)
Das System hat Landmarken in der Wegbeschreibung gut verstanden.	(1)	(2)	(3)	(4)	(5)
Es ist sehr wichtig, dass das System Landmarken versteht.	(1)	(2)	(3)	(4)	(5)
Das System hat Entfernungsangaben gut verstanden.	(1)	(2)	(3)	(4)	(5)
Es ist sehr wichtig, dass das System Entfernungsangaben versteht.	(1)	(2)	(3)	(4)	(5)

Hat die Skizze in dem betreffenden Durchgang wesentlich zu einem besseren Verständnis beigetragen? Ja O Nein O

Verbesserungsvorschläge: ______________________________

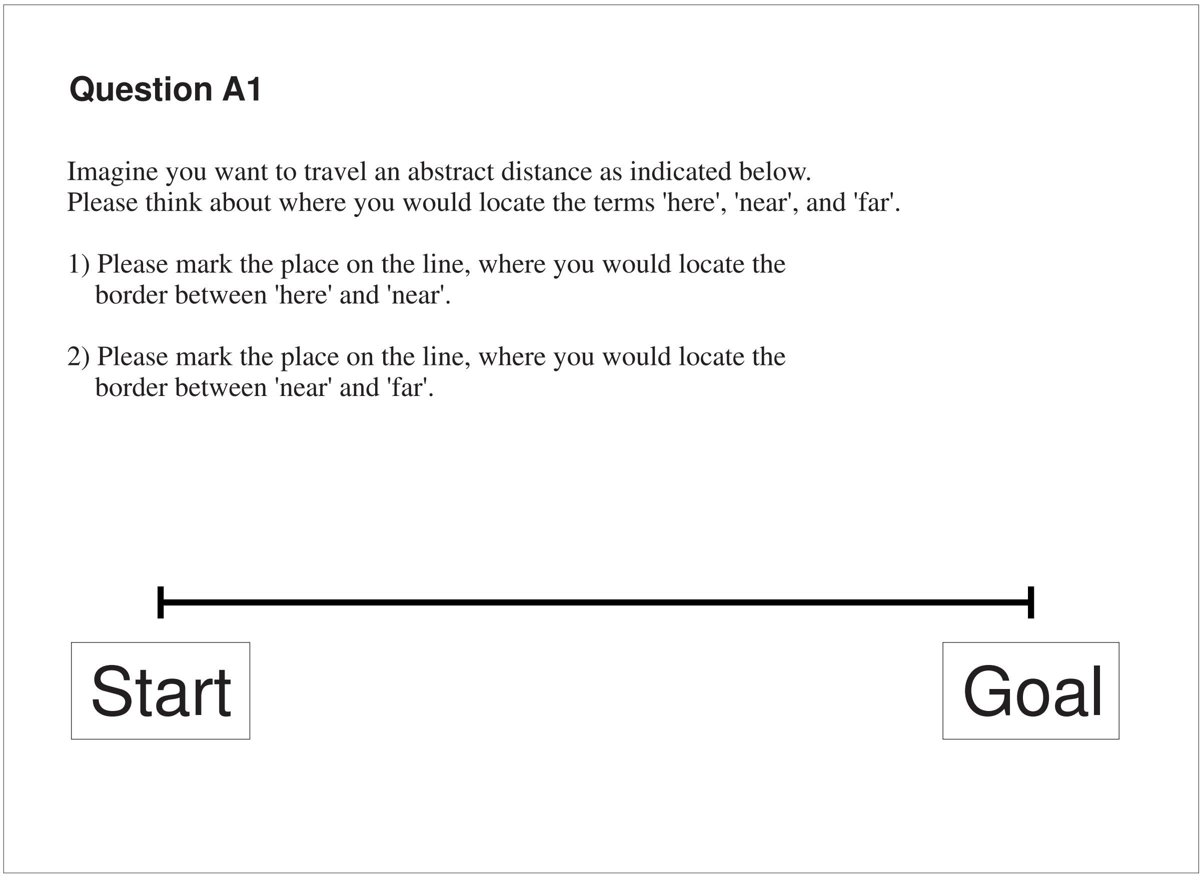
allgemeine Anmerkungen: ______________________________

Fig. C.1: Questionnaire for the subjective assessment of the dialog system.

C.2 Questionnaire for Qualitative Distance Assessment

The questionnaires depicted in Fig. C.2, C.4, and C.3 were applied to identify a contextual dependence for the spatial deictics *'here'*, *'close'*, and *'far'*, as described in Appendix B.2.

The questionnaire in Fig. C.5 was applied to gain data about the distance reaches of the qualitative distance descriptions *'here'*, *'close'*, and *'far'*. The resulting models, based upon the collected data were presented in Section 4.3.3.



Question A1

Imagine you want to travel an abstract distance as indicated below.
Please think about where you would locate the terms 'here', 'near', and 'far'.

1) Please mark the place on the line, where you would locate the border between 'here' and 'near'.

2) Please mark the place on the line, where you would locate the border between 'near' and 'far'.

Start

Goal

Fig. C.2: Questionnaire used to determine the relative borders between spatial deictics for an abstract environment.

Question A2

Imagine you want to walk from building N5 to Marienplatz. Have a look at the map below and think about where you would locate the terms 'here', 'near', and 'far'

1) Please mark the place on the line between N5 and Marienplatz, where you would locate the border between 'here' and 'near'.

2) Please mark the place on the line between N5 and Marienplatz, where you would locate the border between 'near' and 'far'.

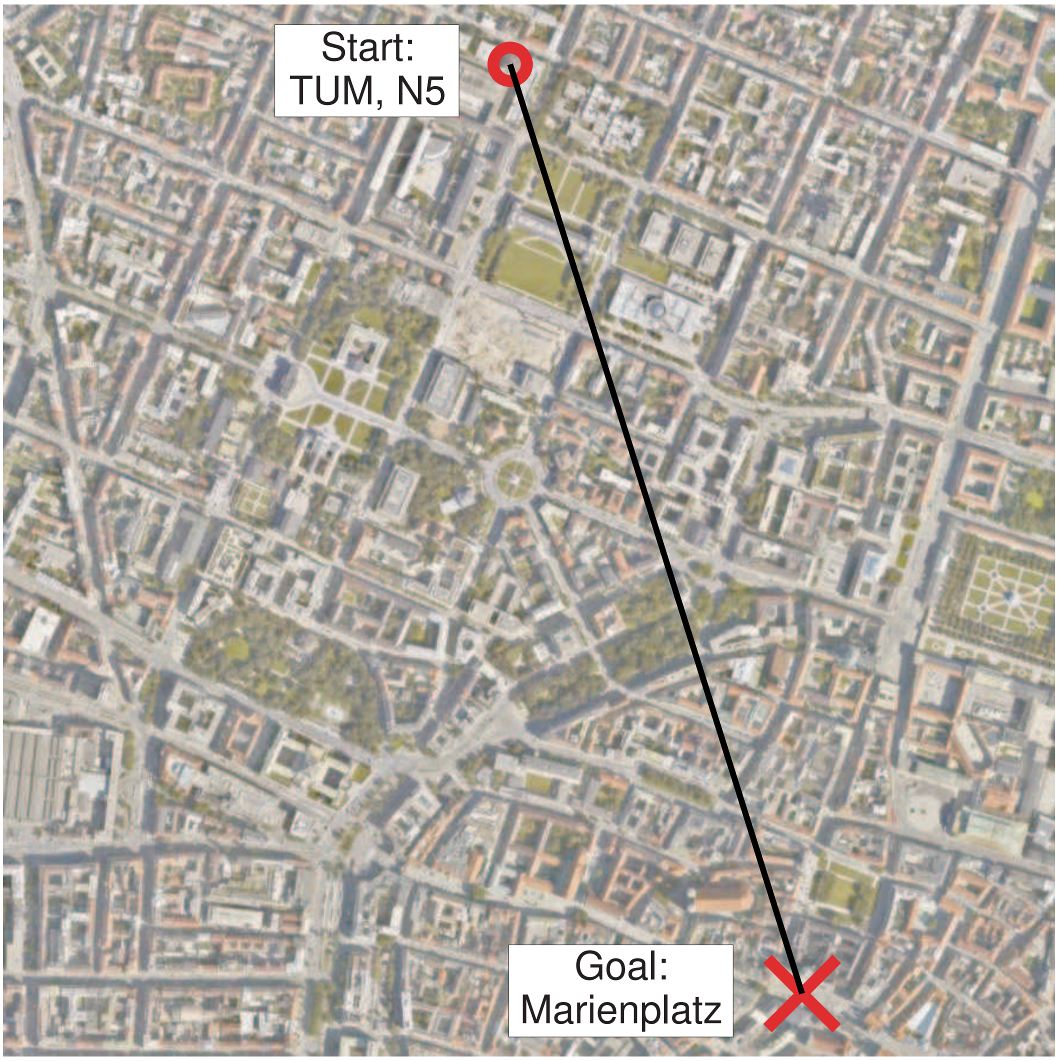


Fig. C.3: Questionnaire used to determine the relative borders between spatial deictics for an urban environment.

Question A3

Imagine you want to drive from Munich to Frankfurt. Have a look at the map below and think about where you would locate the terms 'here', 'near', and 'far'.

1) Please mark the place on the line between Munich and Frankfurt, where you would locate the border between 'here' and 'near'.

2) Please mark the place on the line between Munich and Frankfurt, where you would locate the border between 'near' and 'far'.

Fig. C.4: Questionnaire used to determine the relative borders between spatial deictics for a large scale environment.

Question B

Imagine you are asked for directions by a passer-by in Munich.Think about the terms here , close , and far and how you might use those terms when giving directions.

Now have a look at the map below. The point at which you stand is labeled start, and there are locations (boxes) randomly distributed around the map.
Would you agree that the start point can be labeled as here ? ________

Please label the boxes that in your opinion are characteristic examples for locations that you would denote as here , close , and far , respectively.
(Label a box with H for here , with C for close , and with F for far .)
If you are unsure on how to label a box, just leave it unlabeled.

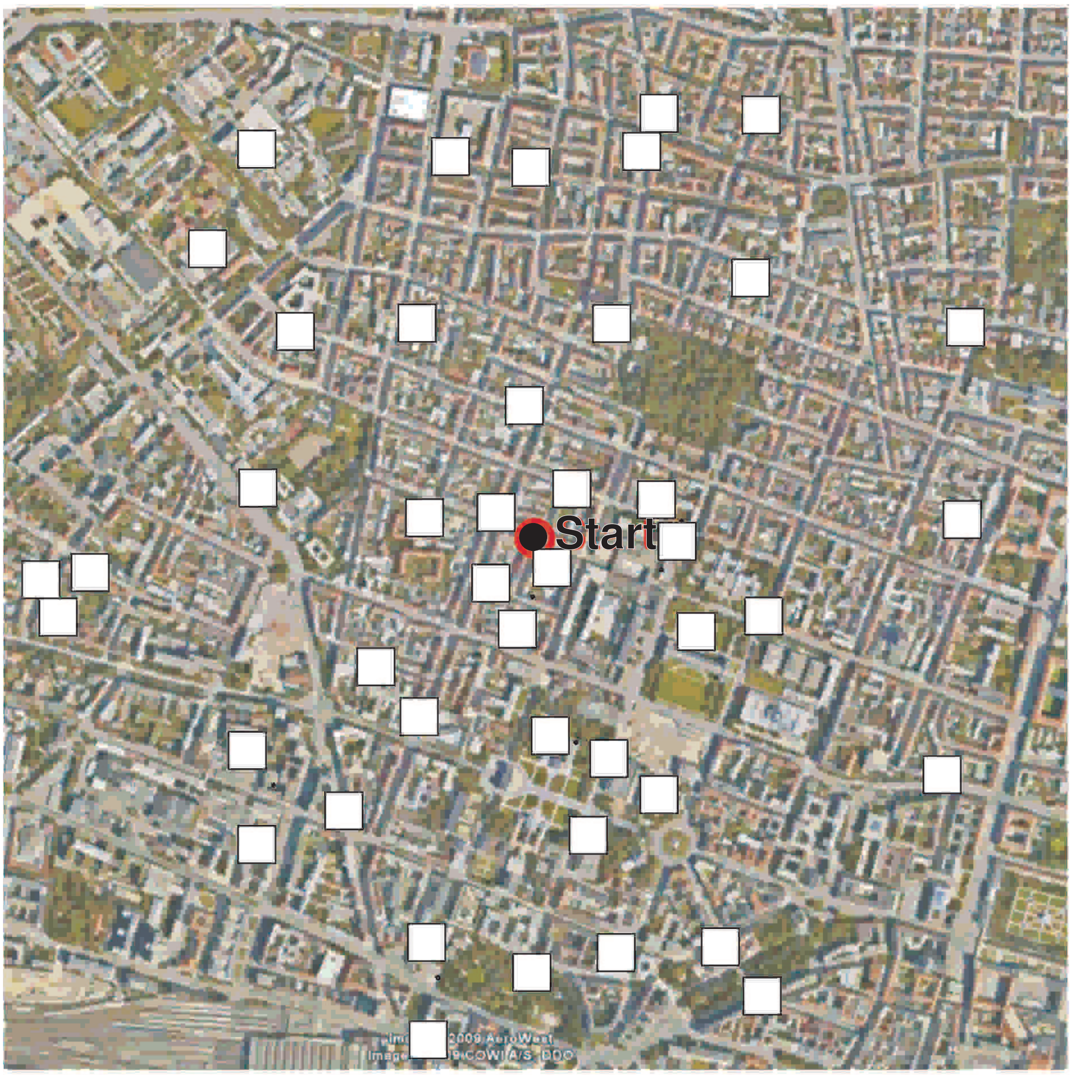


Fig. C.5: Questionnaire used to determine the spatial extends of the spatial deictics *'here'*, *'close'*, and *'far'*, in an urban environment.

Bibliography

[1] Summary report on technology strategy for creating a 'robot society' in the 21st century. Report, 2001. Japan Robot Association.

[2] Robotic visions to 2020 and beyond – the strategic research agenda for robotics in europe. Report, 2009. European Robotics Technology Platform.

[3] S. Anderson and E. Keenan. *Language typology and syntactic description*, volume 3: Grammatical categories and the lexicon, chapter Deixis, pages 259–308. Cambridge University Press, 1985.

[4] G. Apostolico. *Pattern Matching Algorithms*. Oxford Univ Press, 1997.

[5] D. Appleyard. Styles and methods of structuring a city. *Environment and Behavior*, 2(1):100–117, 1970.

[6] H. Asada, M. S. Branicky, C. R. Carignan, H. I. Christensen, R. Fearing, W. Hamel, J. Hollerbach, S. LaValle, M. Mason, B. Nelson, G. Pratt, A. Requicha, B. Ruddy, M. Sitti, G. Sukhatme, R. Tedrake, R. Voyles, and M. Zhang. A roadmap for us robotics: From internet to robotics. Report, 2009.

[7] I. Asimov. *I, Robot*. Voyager, 1968.

[8] H. Asoh, Y. Motomura, F. Asano, I. Hara, S. Hayamizu, N. Vlassis, and B. Kröse. Jijo-2: An office robot that communicates and learns. *Intelligent Systems*, 19(5):46–55, 2001.

[9] J. Bateman and M. Zock. *Oxford Handbook of Computational Linguistics*, chapter Natural language generation, pages 284–304. Oxford University Press, 2002.

[10] T. Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418, 1763.

[11] M. Beetz, J. Bandouch, A. Kirsch, A. Maldonado, A. Müller, and R. B. Rusu. The assistive kitchen - a demonstration scenario for cognitive technical systems. In *Proceedings of the COE Workshop on Human Adaptive Mechatronics*, 2007.

[12] D. Berry. *Statistics: A Bayesian Perspective*. Duxbury, 1995.

[13] I. A. Bolshakov and A. Gelbukh. *Computational Linguistics: Models, Resources, Applications*. Centro de Investigación en Computación, Mexico City, 2004.

[14] D. R. Bradley and D. Vido. Psychophysical functions for perceived and remembered distance. *Perception*, 13(3):315–320, 1984.

[15] J. Brandt and W. Mackavey. Left-right confusion and the perception of bilateral symmetry. *International Journal on Neuroscience*, 12:87–94, 1981.

[16] O. Bratfisch. A further study of the relation between subjective distance and emotional involvement. *Acta Psychologica*, 29:244–255, 1969.

[17] C. Breazeal and B. Scassellati. How to build robots that make friends and influence people. In *Proceedings of the International Conference on Intelligent Robots and Systems*, pages 858–863, Kyongju, Korea, 1999.

[18] R. Briggs. *Image and environment: Cognitive mapping and spatial behavior*, chapter Urban Cognitive Distances, pages 361–88. Aldine Publishing, 1973.

[19] R. Brooks. New approaches to robotics. *Science*, 253:1227–1232, 1991.

[20] D. Brosset, C. Claramunt, and E. Saux. A location and action-based model for route descriptions. In *Lecture Notes in Computer Science: GeoSpatial Semantics*, volume 4853, pages 146–159. Springer, 2007.

[21] M. Buehler, K. Lagnemma, and S. Singh, editors. *Journal of Field Robotics: Special Issue on the 2007 DARPA Urban Challenge, Part I.* Wiley Periodicals, Inc., 2008.

[22] G. Bugmann, E. Klein, S. Lauria, and T. Kyriacou. Corpus-based robotics: A route instruction example. In *Proceedings of Intelligent Autonomous Systems*, pages 96–103, 2004.

[23] K. Bühler. *Theory of language. The representational function of language.* Benjamins, 1934/1990.

[24] W. Burgard, A. B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun. Experiences with an interactive museum tour-guide robot. *Artificial Intelligence*, 114:3–55, 1999.

[25] R. Byrne. Memory for urban geography. *Quarterly Journal of Experimental Psychology*, 31:147–154, 1979.

[26] M. Cadwaller. A methodological analysis of cognitive distance. *Environmental Design Research*, 2:193–199, 1973.

[27] D. Canter and S. Tagg. Distance estimation in cities. *Environment and Behavior*, 7(1):63–63, 1975.

[28] J. Cassell, T. Stocky, T. Bickmore, Y. Gao, Y. Nakano, K. Ryokai, D. Tversky, C. Vaucelle, and H. Vilhjalmsson. Mack: Media lab autonomous conversational kiosk. In *Proceedings of Imagina: Intelligent Autonomous Agents*, 2002.

[29] N. Chater, J. Tenenbaum, and A. Yuille. Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, 10(7):287–291, 2006.

[30] N. Chomsky. *Syntactic Structures.* The Hague: Mouton, 1957.

[31] H. Clark and S. Brennan. *Perspectives on socially shared cognition*, chapter Grounding in communication, pages 127–149. American Psychological Association, 1991.

[32] A. Cohn and J. Renz. *Handbook of Knowledge Representation*, chapter Qualitative spatial reasoning, pages 551–596. Elsevier, 2007.

[33] R. Cole, J. Mariani, H. Uszkoreit, A. Zaenen, and V. Zue. *Survey of the State of the Art in Human Language Technology.* Cambridge University Press, 1997.

[34] R. Cole, B. Wise, and S. van Vuuren. How marni teaches children to read. *Educational Technology*, 47:14–18, 2006.

[35] A. Crompton. Perceived distance in the city as a function of time. *Environment and Behavior*, 38(2):173–182, 2006.

[36] R. Day. Urban distance cognition: review and contribution. *Australian Geographer*, 13(3):193–200, 1976.

[37] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.

[38] C. F. DiSalvo, F. Gemperle, J. Forlizzi, and S. B. Kiesler. All robots are not created equal: the design and perception of humanoid robot heads. In *Proceedings of the Symposium on Designing Interactive Systems*, pages 321–326, 2002.

[39] G. Dissanayake, P. Newman, S. Clark, H. Durrant-Whyte, and M. Csorba. A solution to the simultaneous localization and map building (SLAM) problem. *Transactions on Robotics and Automation*, 17(3):229–241, 2001.

[40] R. Downs and D. Stea. *Image and environment: Cognitive mapping and spatial behavior*, chapter Cognitive maps and spatial behavior: Process and products, pages 8–26. Aldine Publishing, 1973.

[41] S. DuPreez, M. Lall, and S. Sinha. An intelligent web-based voice chat bot. In *Eurocon*, pages 386–39, 2009.

[42] G. Evans, P. Brennan, M. A. Skorpanich, , and D. Held. Cognitive mapping and elderly adults: Verbal and location memory for urban landmarks. *Journal of Gerontology*, 36(4):452–457, 1984.

[43] C. Fillmore. *Universals in linguistic theory*, chapter The case for case, pages 1–88. Holt, Rinehart, and Winston, New York, 1968.

[44] C. Fillmore. *Lectures on Deixis.* University of Chicago Press, 1997.

[45] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots: Concepts, design, and applications. Technical Report CMU-RI-TR-02-29, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 2002.

[46] T. W. Fong, I. Nourbakhsh, R. Ambrose, R. Simmons, A. Schultz, and J. Scholtz. The peer-to-peer human-robot interaction project. In *Proceedings of the AIAA Space 2005*, 2005.

[47] R. Forgus. *Perception: The Basic Process in Cognitive Development*. McGraw-Hill, New York, 1966.

[48] M. E. Foster, T. By, M. Rickert, and A. Knoll. Symmetrical joint action in human-robot dialogue. In *RSS 2006 Workshop on Intuitive Human-Robot Interaction for Getting the Job done*, 2006.

[49] N. Franklin. *The Construction of Cognitive Maps*, chapter Language as a Means of Constructing and Conveying Cognitive Maps, pages 275–295. Springer, 1996.

[50] E. Fricke. Origo, pointing, and speech: the impact of co-speech gestures on linguistic deixis theory. *Gesture*, 2(2):207–226, 2002.

[51] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J. Fernandez-Madrigal, and J. Gonzalez. Multi-hierarchical semantic maps for mobile robotics. In *Proceedings of the Int. Conference on Intelligent Robots and Systems*, pages 2278–2283, 2005.

[52] T. Garling, A. Book, and E. Lindberg. Cognitive mapping of large-scale environments. *Environment and Behavior*, 16(1):3–34, 1984.

[53] S. Gentry and R. Murray-Smith. Haptic dancing: human performance at haptic decoding with a vocabulary. In *Proceedings of the International conference on Systems Man and Cybernetics*, volume 4, pages 3432–3437, Washington D.C., USA, 2003.

[54] A. S. Gilinsky. Perceived size and distance in visual space. *Psychological Review*, 58:460–482, 1951.

[55] R. Golledge, R. Briggs, and D. Demko. The configuration of distances in intra-urban space. In *Proceedings of the Ass. of American Geographers*, pages 60–65, 1969.

[56] R. Golledge and N. Spector. Comprehending the urban environment: theory and practice. *Geographical Analysis*, 14:305–325, 1978.

[57] P. Gould and R. White. *Mental Maps*. Penguin Books, 1974.

[58] D. Griffin. *Foundations of Psychology*, chapter Topographical Orientation, pages 380–392. Wiley, New York, 1948.

[59] H.-M. Gross, H.-J. Böhme, C. Schröter, S. Müller, A. König, C. Martin, M. Merten, and A. Bley. Shopbot: Progress in developing an interactive mobile shopping assistant for everyday use. In *Proceedings of theInternational Conference on Systems, Man and Cybernetics*, 2008.

[60] A. Gruenstein and S. Seneff. Releasing a multimodal dialogue system into the wild: User support mechanisms. In *Proceedings of the Workshop on Discourse and Dialogue*, pages 111–119, 2007.

[61] C. Habel. Wegbeschreibungen in Trier, 1985 / 2008.

[62] A. Hajek. *The Stanford Encyclopedia of Philosophy*, chapter Interpretations of probability. 2003.

[63] E. Hall. A system for the notation of proxemic behaviour. *American Anthropologist*, 65:1003–1026, 1963.

[64] E. Hall. *The Hidden Dimension*. Doubleday, New York, 1966.

[65] Y. Hattori, H. Kozima, K. Komatani, T. Ogata, and H. G.Okuno. Robot gesture generation from environmental sounds using inter-modality mapping. In *Proceedings of the International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, number 123, pages 139–140, Nara, Japan, 2006. Lund University Cognitive Studies.

[66] K. Holyoak and W. Mah. Cognitive reference points in judgments of symbolic magnitude. *cognitive psychology*, 14(3):328–352, 1982.

[67] K. Hone and R. Graham. Towards a tool for the subjective assessment of speech system interfaces (SASSI). *Natural Language Engineering*, 6(3–4):287–303, 2000.

[68] H. Hüttenrauch, K. S. Eklundh, and E. A. T. A. Green. Investigating spatial relationships in human-robot interaction. In *Proceedings of the International Conference on Intelligent Robots and Systems*, pages 5052–5059, 2006.

[69] J. Hutchins. *A new spectrum of translation studies*, chapter Machine translation and computer-based translation tools., pages 13–48. University of Valladolid, 2004.

[70] K. Iwatsuka, K. Yamamoto, and K. Kato. Development of a guide dog system for the blind with character recognition ability. In *Proceedings of the Canadian Conference on Computer and Robot Vision*, pages 401–405, Washington, DC, USA, 2004. IEEE Computer Society.

[71] P. Jia and H. Hu. Head gesture based control of an intelligent wheelchair. In *Proceedings of the Conference of the Chinese Automation and Computing Society in the UK*, pages 85–90, Sheffield, UK, 2005.

[72] K. Jokinen and M. McTear. *Spoken Dialogue Systems*. Morgan & Claypool Publishers, 2008.

[73] B. H. Juang and L. R. Rabiner. *Encyclopedia of Language and Linguistics*, chapter Automatic Speech Recognition - A Brief History of the Technology. Elsevier, second edition, 2005.

[74] H. Jung, C.-N. Seon, J. H. Kim, J. C. Sohn, W.-K. Sung, and D.-I. Park. Information extraction for user's utterance processing on ubiquitous robot companion. *Natural Language Processing and Information Systems*, 3513:337–340, 2005.

[75] D. Jurafsky and J. Martin. *Speech and Language Processing*. Prentice Hall, 2008.

[76] T. Kanade, J. Cohn, and Y.-L. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pages 46–53, 2000.

[77] F. Kaplan. Who is afraid of the humanoid? investigating cultural differences in the acceptance of robots. *Int. Journal of Humanoid Robotics*, 1(3):465–480, 2004.

[78] W. Kennedy, M. Bugajska, M. Marge, W. Adams, B. Fransen, D. Perzanowski, A. Schultz, and G. Trafton. Spatial representation and reasoning for human-robot collaboration. In *Proceedings of the Conference on Artificial Intelligence*, pages 1554–1559, 2007.

[79] S. Kiesler and P. Hinds. Introduction to this special issue on human-robot interaction. *Human-Computer Interaction*, 19(1):1–8, 2004.

[80] W. Klein. Wegauskünfte. *Zeitschrift für Literaturwissenschaft und Linguistik*, 33:9–57, 1979.

[81] W. Klein. *Speech, Place, and Action: Studies in Deixis and Related Topics*, chapter Local Deixis in Route Directions, pages 161–182. New York: Wiley, 1982.

[82] W. Klein. *Spatial Orientation: Theory, Research, and Application*, chapter Deixis and spatial orientation in route directions, pages 283–311. New York: Plenum, 1983.

[83] A. Klippel, C. Dewey, M. Knauff, K.-F. Richter, D. R. Montello, C. Freksa, and E.-A. Loeliger. Direction concepts in wayfinding assistance systems. In *Proceedings of the Workshop on Artificial Intelligence in Mobile Systems*, pages 1–8, 2004.

[84] D. Kortenkamp, E. Huber, and R. P. Bonasso. Recognizing and interpreting gestures on a mobile robot. In *Proceedings of the National Conference on Artificial Intelligence*, volume 2, pages 915–921, Portland, USA, 1996.

[85] S. Kosslyn, H. Pick, and G. Fariello. Cognitive maps in children and men. *Child Development*, 45(3):707–716, 1974.

[86] K. Kosuge, T. Hayashi, Y. Hirata, and R. Tobiyama. Dance partner robot - ms dancer. In *Proceedings of the International Conference on Intelligent Robots and Systems*, volume 4, pages 3459–3464, Las Vegas, USA, 2003.

[87] K. Kosuge and N. Kazamura. Control of a robot handling an object in cooperation with a human. In *Proceedings of the International Workshop on Robot and Human Communication*, pages 142–147, Sendai, Japan, 1997.

[88] K. Kosuge, M. Sato, and N. Kazamura. Mobile robot helper. In *Proceedings of the International Conference of Robotics and Automation*, pages 583–588, 2000.

[89] B. Krieg-Brückner, U. Frese, K. Lüttich, and C. Mandel. Specification of an ontology for route graphs. In *Lecture Notes in Computer Science: Spatial Cognition IV – Reasoning, Action, and Interaction*, pages 390–412. Springer-Verlag, Berlin, 2005.

[90] B. Kuipers. A hierarchy of qualitative representations for space. In *Proceedings of the International Workshop on Qualitative Reasoning about Physical Systems*, pages 113–120. AAAI Press, 1996.

[91] B. Kuipers. The spatial semantic hierarchy. *Artificial Intelligence*, 119:191–233, 2000.

[92] B. Kuipers and Y. Byun. A robust qualitative method for spatial learning in unknown environments. In *Proceedings of the National Conference on Artificial Intelligence*, pages 774–779, 1988.

[93] B. Kuipers and T. Levitt. Navigation and mapping in large-scale space. *AI Magazine*, 9(2):25–43, 1988.

[94] B. Kuipers and A. Stroupe. The AAAI-2002 robot challenge. *Artificial Intelligence Magazine*, 24(1):65–71, 2003.

[95] D. Kulić and E. Croft. Estimating intent for human-robot interaction. In *Proceedings of the International Conference on Advanced Robotics*, pages 810–815, 2003.

[96] D. Kulic and E. Croft. Real-time safety for human-robot interaction. *Robotics and Autonomous Systems*, 54(1):1–12, 2006.

[97] D. Kulić; and E. A. Croft. Safe planning for human-robot interaction. *Journal of Robotic Systems*, 22(7):383–396, 2005.

[98] V. A. Kulyukin, C. Gharpure, J. Nicholson, and G. Osborne. Robot-assisted wayfinding for the visually impaired in structured indoor environments. *Autonomous Robots*, 21(1):29–41, 2006.

[99] Y. Kunii and H. Hashimoto. Tele-handshake using handshake device. In *Proceedings of the International Conference on Industrial Electronics, Control, and Instrumentation*, pages 179–182, 1995.

[100] Y. Kuroki, M. Fujita, T. Ishida, K. Nagasaka, and J. Yamaguchi. A small biped entertainment robot exploring attractive applications. In *Proceedings of the International Conference on Robotics and Automation*, pages 471–476. IEEE, 2003.

[101] S. Lauria, G. Bugmann, T. Kyriacou, and E. Klein. Instruction based learning: how to instruct a personal robot to find HAL. In *Proceedings of the European Workshop on Learning Robots*, pages 15–24, 2001.

[102] T. Lee. Perceived distance as a function of direction in the city. *Environment and Behavior*, 2(1):40–51, 1970.

[103] S. Lemmetty. Review of speech synthesis technology. Master's thesis, Helsinki University of Technology, Department of Electrical and Communications Engineering, 1999.

[104] V. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8):707–710, 1966.

[105] Z. Li, N. Hofemann, J. Fritsch, and G. Sagerer. Hierarchical modeling and recognition of manipulative gesture. In *Proceedings of the Workshop on Modeling People and Human Interaction*, Beijing, China, 2005.

[106] G. Lidoris, F. Rohrmüller, D. Wollherr, and M. Buss. The autonomous city explorer (ace) project - mobile robot navigation in highly populated urban environments. In *Proceedings of the International Conference on Robotics and Automation*, 2009.

[107] R. Lloyd and C. Heivly. Systematic distortions in urban cognitive maps. *Annals of the Association of American Geographers*, 77(2):191–207, 1987.

[108] R. K. Lüneburg. *Mathematical Analysis of Binocular Vision*. Princeton University Press, 1947.

[109] M. Lohse, M. Hanheide, K. Rohlfing, and G. Sagerer. Systemic interaction analysis (sina) in hri. In *Proceedings of the International Conference on Human-Robot Interaction*, pages 93–100, 2009.

[110] K. Lynch. *The Image of the City*. MIT PRess, 1960.

[111] J. Lyons. *Linguistic Semantics: An Introduction*. Cambridge University Press, 1995.

[112] M. MacMahon, B. Stankiewicz, and B. Kuipers. Walk the talk: connecting language, knowledge, and action in route instructions. In *Proceedings of the Conference On Artificial Intelligence*, pages 1475–1482, 2006.

[113] P. Maes, T. Darrell, B. Blumberg, and A. Pentland. The alive system: full-body interaction with autonomous agents. *Computer Animation*, 1:11–18, 1995.

[114] M. Marcus, B. Santorini, and M. Marcinkiewicz. Building a large annotated corpus of english: the penn treebank. *Computational Linguistics*, 19(2):313–330, 1993.

[115] A. Markov. *Dynamic Probabilistic Systems*, volume 1: Markov Chains, chapter Extension of the limit theorems of probability theory to a sum of variables connected in a chain, pages 552–577. Dover Publications, 1971.

[116] T. Matsui, H. Asoh, F. Asano, T. Kurita, I. Hara, Y. Motomura, K. Itou, and J. Fry. Integration of real-world interaction functions on the jijo-2 office robot. *Bulletin of the Electrotechnical Laboratory*, 64:51–60, 2001.

[117] T. McNamara, J. Hardy, and S. Hirtle. Subjective hierarchies in spatial memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15:211–227, 1989.

[118] M. Michalowski, S. Sabanovic, C. DiSalvo, D. B. Font, L. Hiatt, N. Melchior, and R. Simmons. Socially distributed perception: Grace plays social tag at AAAI 2005. *Autonomous Robots*, 22(4):385–397, 2007.

[119] R. Müller, T. Röfer, A. Lankenau, R. Musto, K. Stein, and A. Eisenkolb. *Lecture Notes in Computer Science: Spatial Cognition II – Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications*, chapter Coarse Qualitative Descriptions in Robot Navigation, pages 265–276. Number 1849. Springer Berlin / Heidelberg, 2000.

[120] I. Moar and G. Bower. Inconsistency in spatial knowledge. *Memory & Cognition*, 11:107–113, 83.

[121] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM: A factored solution to simultaneous localization and mapping. In *Proceedings of the Conference on Artificial Intelligence*, 2002.

[122] Y. Morales, E. Takeuchi, A. Carballo, W. Tokunaga, H. Kuniyoshi, A. Aburadani, A. Hirosawa, Y. Nagasaka, Y. Suzuki, and T. Tsubouchi. 1km autonomous robot navigation on outdoor pedestrian paths. In *Proceedings of the International Conference on Intelligent Robots and Systems*, 2008.

[123] H. Moravec and A. Elfes. High resolution maps from wide angle sonar. In *Proceedings of the International Conference on Robotics and Automation*, pages 116–121, 1985.

[124] M. Mori. The uncanny valley. *Energy*, 7(4):33–35, 1970.

[125] R. S. Moyer, D. R. Bradley, M. H. Sorensen, J. C. Whitin, and D. P. Mansfield. Psychophysical functions for perceived and remembered size. *Science*, 200:330–332, 1978.

[126] Q. Mühlbauer, S. Sosnowski, T. Xu, T. Zhang, K. Kühnlenz, and M. Buss. Navigation through urban environments by visual perception and interaction. In *Proceedings of the International Conference on Robotics and Automation*, 2009.

[127] A. Murata. Basic characteristics of human's distance estimation. In *Proceedings of the International Conference on Systems, Man, and Cybernetics*, volume 2, pages 38–43, 1999.

[128] R. R. Murphy. Human-robot interaction in rescue robotics. *Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 34(2):138–153, 2004.

[129] R. R. Murphy, D. Riddle, and E. Rasmussen. Robot-assisted medical reachback: A survey of how medical personnel expect to interact with rescue robots. In *Proceedings of the International Workshop on Robot and Human Interactive Communication*, pages 301–306, 2004.

[130] G. Navarro. A guided tour to approximate string matching. *ACM Computing Surveys*, 33(1):31–88, 2001.

[131] P. Nugues. An overview of speech synthesis and recognition. In *An Introduction to Language Processing with Perl and Prolog*, pages 1–22. Springer, 2006.

[132] J. O'Keefe and L. Nadel. *The Hippocampus as a Cognitive Map*. Oxford University Press, 1978.

[133] Y. Okuno, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita. Providing route directions: Design of robot's utterance, gesture, and timing. In *Proceedings of the International Conference on Human-Robot Interaction*, pages 53–60, 2009.

[134] E. Pacchierotti, H. Christensen, and P. Jensfelt. Design of an office guide robot for social interaction studies. In *Proceedings of the International Conference on Intelligent Robots and Systems*, 2006.

[135] M. Pardowitz, R. Zöllner, S. Knoop, and R. Dillmann. Incremental learning of tasks from user demonstrations, past experience, and vocal comments. *Transactions on Systems, Man and Cybernetics*, 37(2):322–332, 2007.

[136] R. Passini. *Wayfinding in Architecture*. John Wiley & Sons Inc., 1984.

[137] V. Pavlović, R. Sharma, and T. S. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):677–695, 1997.

[138] J. Pearl. Bayesian networks: A model of self-activated memory for evidential reasoning. In *Proceedings of the Conference of the Cognitive Science Society*, pages 329–334, 1985.

[139] J. Piaget and B. Inhelder. *The Child's Conception of Space*. Norton, New York, 1948.

[140] R. Pieraccini and D. Lubensky. Spoken language communication with machines: the long and winding road from research to business. In *Proceedings of the International conference on Innovations in Applied Artificial Intelligence*, pages 6–15. Springer-Verlag, 2005.

[141] G. Pilato, A. Augello, G. Vassallo, and S. Gaglio. Ehebby: An evocative humorist chat-bot. *Mobile Information Systems*, 4(3):165–181, 2008.

[142] J. Pineau, M. Montemerlo, M. Pollack, N. Roy, and S. Thrun. Towards robotic assistants in nursing homes: Challenges and results. *Robotics and Autonomous Systems*, 42(3–4):271–281, 2003.

[143] Pleo. Innvo labs, www.pleoworld.com.

[144] D. R. Proffitt. Distance perception. *Current Directions in Psychological Research*, 15:131–135, 2006.

[145] J. Pustejovsky, P. Hanks, R. Sauri, A. See, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro, and M. Lazo. The timebank corpus. In *Proceedings of Corpus Linguistics*, pages 647–656, 2003.

[146] P. Rani, N. Sarkar, C. A. Smith, and L. D. Kirby. Anxiety detecting robotic system – towards implicit human-robot collaboration. *Robotica*, 22(1):85–95, 2004.

[147] A. Raux, B. Langner, D. Bohus, A. Black, and M. Eskenazi. Let's go public! taking a spoken dialog system to the realworld. In *Proceedings of the International Conference on Speech Communication and Technology*, pages 885–888, 2005.

[148] J. Renz and B. Nebel. *Handbook of Spatial Logics*, chapter Qualitative Spatial Reasoning using Constraint Calculi, pages 161–215. Springer, 2007.

[149] K.-F. Richter. Route direction structure diagrams. In *Proceedings of the International Space Syntax Symposium*, 2005.

[150] Robomow. Friendly robotics, http://www.friendlyrobotics.com/.

[151] RobosapienV2. Wowwee, http://www.robosapienv2online.com/.

[152] D. Rodriguez-Losada, F. Matia, A. Jimenez, R. Galan, and G. Lacey. Implementing map based navigation in guido, the robotic smartwalker. In *Proceedings of the International Conference on Robotics and Automation*, pages 3390–3395, 2005.

[153] Roomba. irobot, http://www.irobot.com.

[154] H. Sacks, E. Schegloff, and G. Jefferson. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50:696–735, 1974.

[155] E. Sadalla and L. Staplin. The perception of traversed distance. *Environment and Behavior*, 12(2):167–182, 1980.

[156] K. Sakita, K. Ogawara, and S. Murakami. Flexible cooperation between human and robot by interpreting human intention from gaze information. In *Proceedings of the International Conference on Intelligent Robots and Systems*, pages 846–851, Sendal, Japan, 2004.

[157] G. Samarasekara, K. Fukahori, and Y. Kubota. Distance cognition at route knowledge level affected by urban vegetation. *SLSAJ Annual Research Journal*, 6:28–33, 2006.

[158] A. Samsonovich and G. Ascoli. Cognitive map dimensions of the human value system extracted from natural language. In B. Goertzel and P. Wang, editors, *Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms*, pages 111–124. IOS Press, Amsterdam, 2007.

[159] S. Sarawagi. *Foundations and Trends in Databases: Information Extraction.* Now Publishers, 2008.

[160] R. Schank. Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology*, 3:552–631, 1972.

[161] M. Schröder. Emotional speech synthesis: A review. In *Proceedings of the European Conference on Speech Communication and Technology*, volume 1, pages 561–564, Aalborg, Denmark, 2001.

[162] M. Schröder and J. Trouvain. The german text-to-speech synthesis system mary: A tool for research, development and teaching. *International Journal of Speech Technology*, 6:365–377, 2003.

[163] P. Sellers. The theory and computation of evolutionary distances: Pattern recognition. *Journal of Algorithms*, 1(4):359–373, 1980.

[164] H. Shi and B. Krieg-Brückner. Modelling route instructions for robust human-robot interaction on navigation tasks. *International Journal of Software and Informatics*, 2:33–60, 2008.

[165] A. Siegel and S. White. *The development of spatial representations of large-scale environments*, volume 10. Academic Press, 1975.

[166] A. Singhal. Modern information retrieval: A brief overview. In *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, volume 24, pages 35–43. IEEE, 2001.

[167] M. Skubic, D. Perzanowski, S. Blisard, A. Schultz, W. Adams, M. Bugajska, and D.Brock. Spatial language for human-robot dialogs. *Transactions on Systems, Man, and Cybernetics*, 34(2):154–167, 2004.

[168] R. Smith and R. Hipp. *Spoken Natural Language Dialog Systems: A Practical Approach*. Oxford Univ Press, 1994.

[169] S. Sosnowski, A. Bittermann, K. Kühnlenz, and M. Buss. Design and evaluation of emotion-display eddie. In *Proceedings of the International Conference on Robotics and Intelligent Systems*, pages 3113–3118, Bejing, China, 2006.

[170] R. Sparrow. The march of the robot dogs. *Ethics and Information Technology*, 4(4):305–318, 2002.

[171] T. Spexard, S. Li, B. Wrede, J. Fritsch, G. Sagerer, O. Booij, Z. Zivkovic, B. Terwijn, and B. Kröse. Biron, where are you? Enabling a robot to learn new places in a real home environment by integrating spoken dialog and visual localization. In *Proceedings of the International Conference on Intelligent Robots and Systems*, pages 943–940, 2006.

[172] A. Stevens and P. Coupe. Distortions in judged spatial relations. *Cognitive Psychology*, 10:422–437, 1978.

[173] S. Stevens. On the psychophysical law. *Psychological Review*, 64(3):153–181, 1957.

[174] H. Strik, A. Russel, H. V. D. Heuvel, C. Cucchiarini, and L. Boves. A spoken dialog system for the dutch public transport information service. *International Journal of Speech Technology*, 2:121–131, 1997.

[175] O. Sugiyama, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita. Natural deictic communication with humanoid robots. In *Proceedings of the International Conference on Intelligent Robots and Systems*, pages 1441–1448, 2007.

[176] W. Sung, M. Song, and K. won Chung. Applying sasang typology theory to robot appearance design. In *Proceedings of the International Conference on Robot & Human Interactive Communication*, pages 1066–1071, 2007.

[177] D. Syrdal, K. Dautenhahn, M. Walters, and K. Koay. Sharing spaces with robots in a home scenario – anthropomorphic attributions and their effect on proxemic expectations and evaluations in a live hri trial. In *Proceedings of the Symposium on AI in Eldercare: New Solutions to Old Problems*, 2008.

[178] D. S. Syrdal, K. L. Koay, M. Walters, and K. Dautenhahn. A personalized robot companion? – the role of individual differences on spatial preferences in hri scenarios. In *Proceedings of the International Conference on Robot & Human Interactive Communication*, pages 1143–1148, 2007.

[179] A. Takanishi, Y. Ogura, and K. Itoh. *Robotics Research*, chapter Some Issues in Humanoid Robot Design, pages 357–372. Springer Berlin Heidelberg, 2007.

[180] P. Thorndyke and B. Hayes-Roth. Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology*, 14(4):560–589, 1982.

[181] S. Thrun, M. Bennewitz, W. Burgard, A. B. Cremers, F. Dellaert, D. Fox, D. Hahnel, C. R. Rosenberg, N. Roy, J. Schulte, and D. Schulz. MINERVA: A tour-guide robot that learns. In *Proceedings of the German Conference on Artificial Intelligence: Advances in Artificial Intelligence*, pages 14–26, 1999.

[182] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. MIT Press, 2005.

[183] S. Thrun, J. Gutmann, D. Fox, W. Burgard, and B. J. Kuipers. Integrating topological and metric maps for mobile robot navigation: A statistical approach. In *Proceedings of the Conference on Artificial Intelligence*, pages 989–995, 2008.

[184] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, K. Lau, C. Oakley, M. Palatucci, V. Pratt, P. Stang, S. Strohband, C. Dupont, L.-E. Jendrossek, C. Koelen, C. Markey, C. Rummel, J. van Niekerk, E. Jensen, P. Alessandrini, G. Bradski, B. Davies, S. Ettinger, A. Kaehler, A. Nefian, and P. Mahoney. Stanley: The robot that won the DARPA grand challenge: Research articles. *Journal of Robotic Systems*, 23:661–692, 2006.

[185] E. Tolman. Cognitive maps in rats and men. *Psychological Review*, 55:189–208, 1948.

[186] V. Traver, A. del Pobil, and M. Perez-Francisco. Making service robots human-safe. In *Proceedings of the International Conference on Intelligent Robots and Systems*, volume 1, pages 696–701, Takamatsu, Japan, 2000.

[187] C. Trowbridge. On fundamental methods of orientation and 'imaginary maps'. *Science*, 38(990):888–897, 1913.

[188] B. Tversky. Distortions in memory for maps. *Cognitive Psychology*, 13:407–433, 1981.

[189] I. Ulrich and J. Borenstein. The GuideCane - applying mobile robot technologies to assist the visually impaired. *Transactions on Systems, Man, and Cybernetics*, 31(2):131–136, 2001.

[190] W. Vapnik. *The Nature of Statistical Learning Theory*. Springer Verlag, 1995.

[191] S. Vasudevan, S. Gächter, M. Berger, and R. Siegwart. Cognitive maps for mobile robots – an object based approach. *Robotics and Autonomous Systems*, 55(5):359–371, 2007.

[192] T. Vintsyuk. Speech discrimination by dynamic programming. *Cybernetics*, 4:52–58, 1968.

[193] K. Wada and T. Shibata. Robot therapy in a care house - its sociopsychological and physiological effects on the residents. In *Proceedings of the International Conference on Robotics and Automation*, pages 3966–3971, 2006.

[194] Y. Wakabayashi. *Geographical Reports of Tokyo Metropolitan University*, volume 29, chapter Spatial analysis of cognitive maps, pages 57–102. 1994.

[195] Wakamaru. Mitsubishi heavy industries ltd., www.mhi.co.jp/en/products/detail/wakamaru.html.

[196] M. Walker, R. Passonneau, and J. Boland. Quantitative and qualitative evaluation of darpa communicator spoken dialogue systems. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 515–522, 2001.

[197] Y. Wei, E. Brunskill, T. Kollar, and N. Roy. Where to go: Interpreting natural directions using global inference. In *Proceedings of the International Conference on Robotics and Automation*, pages 3761–3767, 2009.

[198] J. Weizenbaum. Eliza - a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966.

[199] S. Werner, B. Krieg-Brückner, and T. Herrmann. *Lecture Notes in Computer Science: Spatial Cognition II – Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications*, volume 1849, chapter Modelling Navigational Knowledge by Route Graphs, pages 295–316. Springer-Verlag, Berlin, 2000.

[200] I. Werry, K. Dautenhahn, B. Ogden, and W. Harwin. Investigating a robot as a therapy partner for children with autism. In *Proceedings of the European Conference for the Advancement of Assistgive Technology*. Ljubljana, Slovenia, 2001.

[201] D. Wolter. *Spatial Representation and Reasoning for Robot Mapping.* Springer, Berlin, 2008.

[202] W. A. Woods. Understanding subsumption and taxonomy: A framework for progress. In J. F. Sowa, editor, *Principles of Semantic Networks: Explorations in the Representation of Knowledge*, pages 45–94. Kaufmann, San Mateo, 1991.

[203] D. Wunderlich. *Linguistische Berichte*, volume 53, chapter Wie analysiert man Gespräche? Beispiel: Wegauskünfte, pages 41–76. Helmut Buske Verlag, 1978.

[204] S. Yamada and T. Komatsu. Designing simple and effective expression of robot's primitive minds to a human. In *Proceedings of the International Conference on Intelligent Robots and Systems*, pages 2614–2619, 2006.

[205] Z. Yang and D. Purves. A statistical explanation of visual space. *Nature Neuroscience*, 6:632–640, 2003.

[206] B.-J. You, M. Hwangbo, S.-O. Lee, S.-R. Oh, Y. D. Kwon, and S. Lim. Development of a home service robot 'issac'. In *Proceedings of the International Conference on Intelligent Robots and Systems*, pages 2630–2635, 2003.

[207] M. You, C. Chen, J. Bu, J. Liu, and J. Tao. A hierarchical framework for speech emotion recognition. In *Proceedings of the International Symposium on Industrial Electronics*, volume 1, pages 515–519, 2006.

[208] F. Yus. Misunderstandings and explicit/implicit communication. *Pragmatics*, 9(4):487–517, 1999.

[209] H. Zender, G. Kruijff, and I. Kruijff-Korbayová. A situated context model for resolution and generation of referring expressions. In *Proceedings of the European Workshop on Natural Language Generation*, pages 126–129, 2009.

[210] H. Zender, O. M. Mozos, P. Jensfelt, G.-J. Kruijff, and W. Burgard. Conceptual spatial representations for indoor mobile robots. *Robotics and Autonomous Systems*, 56(6):493–502, 2008.

Author's Publications

[211] A. Bauer and M. Buss. Acquiring route graphs from human instructions. In *Poster Proceedings International Conference on Spatial Cognition*, pages 45–48, 2008.

[212] A. Bauer, B. Gonsior, D. Wollherr, and M. Buss. Rules for human-robot interaction based on principles from linguistics - asking for directions. In *Proceedings of the AISB Symposium on New Frontiers in Human-Robot Interaction*, pages 24–30, 2009.

[213] A. Bauer, B. Gonsior, D. Wollherr, and M. Buss. Dialog system for human-robot interaction for asking for directions guidelines for robots proactively obtaining information. *Computational Linguistics*, under review.

[214] A. Bauer, K. Klasing, G. Lidoris, Q. Mühlbauer, F. Rohrmüller, S. Sosnowski, T. Xu, K. Kühnlenz, D. Wollherr, and M. Buss. The autonomous city explorer: Towards natural human-robot interaction in urban environments. *International Journal of Social Robotics*, 1(2):127–140, 2009.

[215] A. Bauer, K. Klasing, T. Xu, S. Sosnowski, G. Lidoris, Q. Mühlbauer, T. Zhang, F. Rohrmüller, D. Wollherr, K. Kühnlenz, and M. Buss. The autonomous city explorer (video). In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2009. **Most Innovative Video Award**.

[216] A. Bauer, K. Klasing, T. Xu, S. Sosnowski, G. Lidoris, Q.Mühlbauer, T. Zhang, F. Rohrmüller, D. Wollherr, K. Kühnlenz, and M. Buss. The autonomous city explorer project (video). In *Proceedings of the International Conference on Robotics and Automation*, pages 1595–1596, 2009.

[217] A. Bauer, D. Wollherr, and M. Buss. Human-robot collaboration: A survey. *International Journal of Humanoid Robotics*, 5(1):47 – 66, 2008.

[218] A. Bauer, D. Wollherr, and M. Buss. Information retrieval system for human-robot communication - asking for directions. In *Proceedings of the International Conference on Robotics and Automation*, pages 4150–4155, 2009.

[219] A. Bauer, D. Wollherr, and M. Buss. Route information retrieval through hri (poster). AISB Convention, 2009. **Best Student Poster**.

[220] A. Bauer, D. Wollherr, and M. Buss. Probabilistic models for route description information by humans. *SMC*, under review.

[221] A. Bauer, D. Wollherr, and M. Buss. Simultaneous reasoning and mapping: Building reasonable route belief from human route directions. *AI*, under review.

[222] B. Gonsior and A. Bauer. Wegbeschreibungen in München, 2010. www.lsr.ei.tum.de/fileadmin/publications/MuenchenKorpus_2009.pdf.

[223] K. Klasing, G. Lidoris, A. Bauer, F. Rohrmüller, D. Wollherr, and M. Buss. The autonomous city explorer: Towards semantic navigation in urban environments. In *Proceedings of the International Workshop on Cognition for Technical Systems*, 2008.

[224] G. Lidoris, K. Klasing, A. Bauer, T. Xu, K. Kühnlenz, D. Wollherr, and M. Buss. The autonomous city explorer project: Aims and system overview. In *Proceedings of the Int. Conference on Intelligent Robots and Systems*, pages 560–565, 2007.

[225] Q. Mühlbauer, T. Xu, A. Bauer, K. Klaasing, G. Lidoris, F. Rohrmüller, S. Sosnowski, K. Kühnlenz, D. Wollherr, and M. Buss. Wenn Roboter nach dem Weg fragen. *at – Automatisierungstechnik*, 58 11:647–656, 2010.

[226] A. Weiss, J. Igelsböck, M. Tscheligi, A. Bauer, K. Kühnlenz, D. Wollherr, and M. Buss. Robots asking for directions: The willingness of passers-by to support robots. In *Proceedings of the International Conference on Human-Robot Interaction*, pages 23–30, 2010.