Multimodal Shared-Control Interaction for Mobile Robots in AAL Environments

Cui Jian

Kumulative Dissertation
zur Erlangung des Grades eines Doktors der Ingenieurwissenschaften
– Dr.-Ing. –

Vorgelegt im Fachbereich 3 (Mathematik und Informatik)
Universität Bremen

September 2013

Datum des Promotionskolloquiums: 18. Dezember 2013

Gutachter

Prof. Dr. Bernd Krieg-Brückner (Universität Bremen)
Prof. Dr. John Bateman (Universität Bremen)

Abstract

This dissertation investigates the design, development and implementation of cognitively adequate, safe and robust, spatially-related, multimodal interaction between human operators and mobile robots in Ambient Assisted Living environments both from the theoretical and practical perspectives. By focusing on different aspects of the concept *Interaction*, the essential contribution of this dissertation is divided into three main research packages; namely, *Formal Interaction*, *Spatial Interaction* and *Multimodal Interaction in AAL*. As the principle package, in *Formal Interaction*, research effort is dedicated to developing a formal language based interaction modelling and management solution process and a unified dialogue modelling approach. This package aims to enable a robust, flexible, and context-sensitive, yet formally controllable and tractable interaction. This type of interaction can be used to support the interaction management of any complex interactive systems, including the ones covered in the other two research packages. In the second research package, *Spatial Interaction*, a general qualitative spatial knowledge based multi-level conceptual model is developed and proposed. The goal is to support a spatially-related interaction in human-robot collaborative navigation. With a model-based computational framework, the proposed conceptual model has been implemented and integrated into a practical interactive system which has been evaluated by empirical studies. It has been particularly tested with respect to a set of high-level and model-based conceptual strategies for resolving the frequent spatially-related communication problems in human-robot interaction. Last but not least, in *Multimodal Interaction in AAL*, attention is drawn to design, development and implementation of multimodal interaction for elderly persons. In this elderly-friendly scenario, ageing-related characteristics are carefully considered for an effective and efficient interaction. Moreover, a standard model based empirical framework for evaluating multimodal interaction is provided. This framework was especially applied to evaluate a minutely developed and systematically improved elderly-friendly multimodal interactive system through a series of empirical studies with groups of elderly persons.

Zusammenfassung

Die vorliegende Doktorarbeit untersucht die Konzeption, Entwicklung und Umsetzung von kognitiv adäquater, sicherer und robuster raumbezogener multimodaler Interaktion zwischen Menschen und mobilen Robotersystemen im Rahmen des altersgerechten umgebungsunterstützten Lebens (AAL), aus theoretischer und praktischer Perspektive. Entsprechend den unterschiedlichen Aspekten des zentralen Konzeptes Interaktion, ist der wesentliche Beitrag dieser Arbeit in drei Forschungspakete aufgeteilt, nämlich *Formale Interaktion*, *Räumliche Interaktion* und *Multimodale Interaktion im Kontext von AAL*. Im grundlegenden Paket, *Formale Interaktion*, besteht ein Großteil der Forschungsarbeiten in der Entwicklung eines Lösungsprozesses, der auf einer formalen Sprache basiert und in Modellierung und Management allgemeiner Interaktion eingesetzt werden kann, sowie eines generellen hybriden Ansatzes zur Dialog-Modellierung. Dieses Paket hat das Ziel, eine robuste, flexible und kontext-sensitive, zugleich formal steuerbare und verfolgbare Interaktion zu ermöglichen, die dann dazu verwendet werden kann, Interaktionsmanagement von komplexen interaktiven Systemen zu unterstützen, einschließlich der in den beiden anderen Forschungspaketen abgedeckten Systeme. In dem zweiten Forschungspaket, *Räumliche Interaktion*, wird ein auf qualitativem räumlichen Wissen basierendes, allgemeines konzeptionelles Mehrebenmodell entwickelt und vorgeschlagen. Das Ziel ist es, eine raumbezogene Interaktion in kooperativer Navigation von Mensch und Roboter zu unterstützen. Das konzeptionelle Modell wurde mit einem modell-basierten Rahmenwerk implementiert und in ein praktisches interaktives System integriert, das dann durch empirische Experimente evaluiert wurde. Dies wurde vor allem im Hinblick auf eine Reihe von modell-basierten konzeptionellen Strategien auf hoher Ebene getestet, die zur Bewältigung der häufigen raumbezogenen Kommunikationsprobleme in Mensch-Roboter-Interaktion verwendet werden. Die Forschungsarbeiten im Paket *Multimodale Interaktion in Umgebungsunterstütztem Leben* konzentrieren sich auf Entwurf, Entwicklung und Implementierung von multimodaler Interaktion für ältere Menschen. Dabei wurden altersbedingte Eigenschaften für eine effektive und effiziente Interaktion in altersgerechter Umgebung sorgfältig betrachtet. Darüber hinaus wurde ein empirisches Rahmenwerk auf der Grundlage des Standard-Modells für die Bewertung multimodaler Interaktion entwickelt. Dieses Rahmenwerk wurde dann speziell angewendet, um ein umfassend entwickeltes und systematisch verbessertes, altersgerechtes multimodales interaktives System durch eine Reihe von empirischen Experimenten mit Gruppen von älteren Menschen zu evaluieren.

Contents

1. General Introduction **1**
1.1. General Overview 1
1.2. Related Work 3
1.2.1. Interaction Management 3
1.2.2. Interaction in Spatially-related Applications 4
1.2.3. Interaction in AAL 5
1.3. Contributions of this Work 6

2. Formal Interaction **9**
2.1. The Formal Language based Dialogue Modelling and Management 9
2.1.1. The Formal Language based Interaction and Modelling Solution Process 10
2.1.2. The Formal Dialogue Development Framework 11
2.2. the Unified Dialogue Modelling and Management Approach 12
2.3. Contribution of the Corresponding Publications 14
2.4. Possible Future Work 14

3. Spatial Interaction **17**
3.1. QSBM: A general Qualitative Spatial Beliefs Model 17
3.2. A DCC-based QSBM 19
3.2.1. The Conceptual Level 19
3.2.2. The Application Level 20
3.2.3. The Strategy Level 22
3.3. SimSpace: A Computational Framework to Support QSBM 25
3.4. Empirical Studies of QSBM based Spatial Interaction 26
3.5. Contribution of the Corresponding Publications 27
3.6. Possible Future Work 28

4. Multimodal Interaction in AAL **29**
4.1. Foundation of Design and Development of Multimodal Interaction for Elderly Persons 29
4.1.1. Design Guidelines of Multimodal Interaction for Elderly Persons 29
4.1.2. The Unified Dialogue Model 31
4.2. MIGSEP: the Multimodal Interactive Guidance System for Elderly Persons . . 32
4.3. Empirical Evaluation of MIGSEP 34
4.4. Contribution of the Corresponding Publications 37
4.5. Possible Future Work 38

Full List of Publications by the Author **42**

Bibliography **51**

A. Accumulated Publications **53**
A.1. SimSpace: A Tool to Interpret Route Instructions with Qualitative Spatial Knowledge . . . 55
A.2. Qualitative Spatial Modelling of Human Route Instructions to Mobile Robots . 57
A.3. Deep Reasoning in Clarification Dialogues with Mobile Robots . . . 63
A.4. Evaluation of a Unified Dialogue Model for Human-Computer Interaction . . . 69
A.5. Towards Effective, Efficient and Elderly-friendly Multimodal Interaction . . . 83
A.6. Evaluating A Spoken Language Interface of A Multimodal Interactive Guidance System for Elderly Persons . . . 91
A.7. Touch and Speech: Multimodal Interaction for Elderly Persons . . . 101
A.8. Better Choice? Combining Speech And Touch In Multimodal Interaction For Elderly Persons . . . 117
A.9. Resolving Conceptual Mode Confusion with Qualitative Spatial Knowledge in Human-Robot Interaction . . . 127
A.10. Modality Preference in Multimodal Interaction for Elderly Persons . . . 145
A.11. A Conceptual Model for Human-Robot Collaborative Spatial Navigation . . . 161

Chapter 1.

General Introduction

In the context of Ambient Assisted Living (AAL), intelligent mobile robots, also known as (semi)automated mobile systems that are capable of navigating human operators through complex spatial environments, are gaining increasing interest in the areas of both academic and industrial research (e.g. see [Lankenau and Röfer, 2000], [Röfer et al., 2009]). Various so-called intelligent assistants concerned with different behaviours of controlling and navigating mobile robots have been developed and evaluated ([17]), some of them can assist human operators to avoid obstacles by taking control themselves if necessary ([Lankenau and Röfer, 2001]), some can go along preassigned routes or to predefined locations fully autonomously ([Röfer and Lankenau, 2002]). Since the mobile robots are collaboratively controlled by these intelligent systems and the human operators usually only have a naive theory about the systems, there inevitably arise problems when the human operators and the mobile robots interact with each other.
Motivated by the need for bridging the communication gap between the human operators and the mobile robots, the research work[1] reported in this dissertation has been focusing on designing, developing and implementing cognitively adequate, safe and robust, spatially-related interaction between human operators and mobile robots in AAL environments.
This chapter will first give a general overview of the research work done by the author, then characterize some of the related research efforts also concerned with the aspects being in the focus of this dissertation, and end the introduction by briefly describing the major contributions of the reported work in the relevant areas.

1.1. General Overview

As illustrated in figure 1.1, by focusing on three different aspects of the concept *Interaction*, the essential contribution of this dissertation is divided into three major research packages: the principle package *Formal Interaction* and two domain-dependent packages *Spatial Interaction* and *Multimodal Interaction in AAL*, each of which is given a brief introduction below.

Formal Interaction has been focusing on developing robust, flexible, context-sensitive yet formally controllable and tractable interaction. This research package consists of two

[1]This work has been funded by the German Research Foundation (DFG) in context of the Sonderforschungsbereich/Transregio 8 *Spatial Cognition*, projects I3-[SharC] and I5-[DiaSpace], as well as the German Research Center for Artificial Intelligence (Deutsches Forschungszentrum für Künstliche Intelligenz, DFKI)

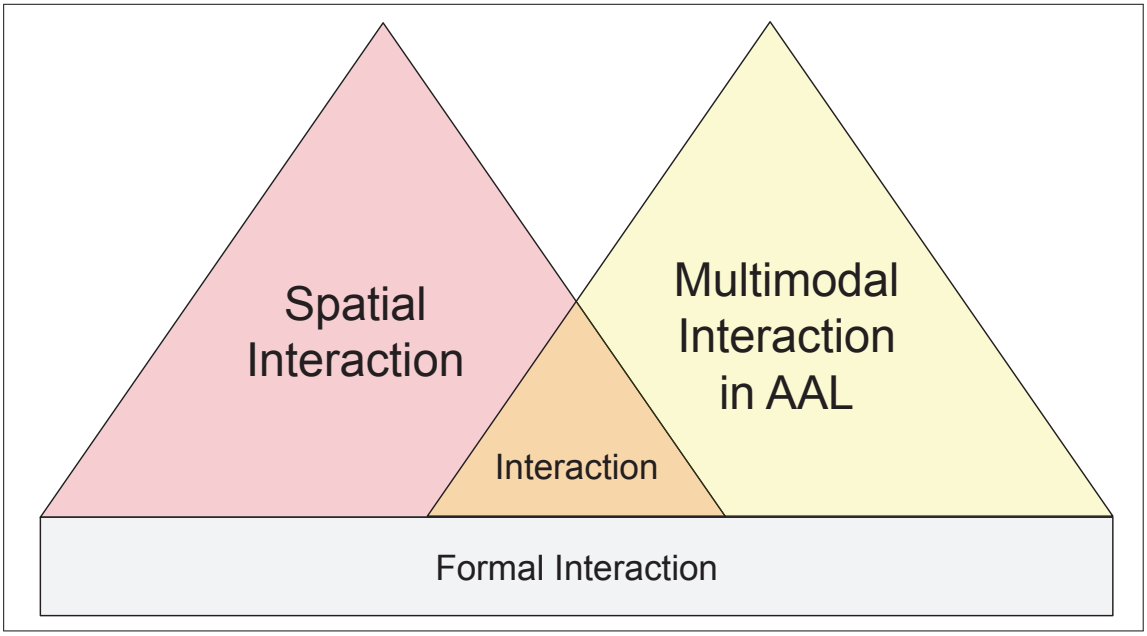


Figure 1.1.: The overview of the reported research work.

core aspects: a) a solution process highlighting a formal language based dialogue modelling and management; and b) a unified dialogue modelling approach to enable a flexible and context-sensitive yet easily manageable interaction. In this package, both theoretical and practical interaction models and frameworks have been delivered for developing and implementing formal dialogue managers in complex interactive systems. Furthermore, they were also used to support the other two research packages in this dissertation.

Spatial Interaction has been aiming at the area of human-robot collaborative navigation within complex spatial environments. Specifically, this research package has been elaborating on problems about how to enable human operators to interact with mobile robots to go from one location to another, while assisting in negotiating the possible communication problems that occur frequently during the interaction. A qualitative spatial knowledge based multi-level conceptual model is developed and proposed. Furthermore, with a model-based computational framework, the conceptual model has been implemented and integrated into a practical interactive system, which was evaluated by empirical studies with respect to a set of high-level and model-based conceptual strategies.

Multimodal Interaction in AAL has been concentrating on effective, efficient and elderly-friendly multimodal interaction in the context of Ambient Assisted Living. This research package consists of two parts: a) the design, development and implementation of multimodal interaction for elderly persons while carefully considering age-related characteristics; and b) the general-framework-based empirical evaluation of a minutely developed and systematically improved elderly-friendly multimodal interactive guidance system. This focused on the effectiveness, efficiency and user-acceptance of the entire system as well as the different input modalities, and was supported by a series

of empirical studies with several groups of elderly persons.

1.2. Related Work

The three research packages have addressed work in the areas of interaction management, interaction in spatially-related applications and interaction in Ambient Assisted Living. Therefore, this section gives an introduction to other research work in the areas in the focus of the dissertation.

1.2.1. Interaction Management

In the context of natural language processing, interaction management, also known as dialogue management, manages the controlling process of an interactive system, which accepts input from the interaction partner, decides upon the next system actions according to the maintained interaction context, and outputs the system responses at a concept level. According to how the interaction flow is controlled, two classic approaches have been proposed for interaction management: structure-based and the principle-based interaction management.

Typical examples of structure-based interaction management can be found in the systems presented in [Peckham, 1993, McTear, 1998, Lamel et al., 1999], where the interaction involved usually has clearly defined structures and goals, and therefore can be modeled as a finite state transition network to enable a straightforward and effective development of interaction management. However, the finite state transition network based management can only control an inflexible interaction flow.

To overcome these problems, other research has been investigating the principle-based interaction management. E.g., the interaction management presented in [Chu-Carroll, 1999, Seneff and Polifroni, 2000, Zue et al., 2000] all shared one principle in common, that the context of the interaction is fixed and can be represented as a set of slots that need to be filled during the interaction, either the departure time of a train for a ticket reservation system, or the goal of a route to be planned, and so on. The interaction is not controlled by a predefined structure, but with a frame-based mechanism, where only if the slots of the frame are filled, specific tasks can be performed. However, instead of dealing with only limited predefined tasks, [Larsson and Traum, 2000, Traum and Larsson, 2003] proposed another principle-based interaction management method: the information state update approach, which manages an interaction flow by defining a set of informational components for functional aspects of interaction such as Question under discussion, common ground, etc., and a set of update rules and update strategies for managing the interaction context, such that the interaction is being managed from the perspective of a human. This approach is now widely used in many interactive systems (e.g., [Lemon and Liu, 2006, Varges et al., 2008], etc.) for its ability to deal with flexible and context-sensitive interaction.

Furthermore, the research community of interaction management was also focusing on the development of stochastic dialogue modeling using reinforcement learning (RL), where statistical data based dialogue modelling was applied to dynamically allow changes to the dialogue strategy (e.g., [Lecoeuche, 2001, Li et al., 2009, Pietquin et al., 2011]). However, this approach is still not that mature to be applied in developing practical interactive systems due

to the requirement of a sufficiently large number of natural language dialogue corpora for the correspondingly large state space and policy space.

1.2.2. Interaction in Spatially-related Applications

Since the research of interaction in spatially-related applications usually involves many aspects in human-computer interaction, cognitive science and robotics, much effort has also been invested in different related aspects from different perspectives.
Some research has been concentrating on the most straightforward way of collecting and analyzing empirical corpora concerned with human-robot interaction e.g. by using natural language route instructions in spatial navigation (e.g., [Bugmann et al., 2004, Koulouri and Lauria, 2009, Shi and Tenbrink, 2009]), which also specified the conceptual as well as the spatially-related difficulties for either human operators to provide route instructions, or mobile robots to process route instructions. Meanwhile, according to empirical findings, effort has also been put into studying the relationship between language and the functional properties of spatial environments (e.g., [Hirtle, 2008]), as well as the natural language route directions or instructions used during the interaction (e.g., [Kollar et al., 2010, Pappu and Rudnicky, 2012]); some even tried to build the conceptual mapping between natural language route instructions and mobile robot executable procedures ([Lauria et al., 2002]).
Apart from the research based on empirical data and natural language, considerable focus has also been placed on how to appropriately represent spatial knowledge to support the spatially-related application. For example, in mobile robotics, metrical spatial data has been related with semantic information based on probabilistic models to resolve the object-recognition based spatial localization problems (e.g., [Galindo et al., 2005, Vasudevan et al., 2007]). Meanwhile, in cognitive science, as a classic conceptual model, [Werner et al., 2000] proposed the Route Graph, which provided a simple, abstract, yet powerful formalism to serve as the basis of complex navigational knowledge and support route-based navigation. This model was further improved and adapted with respect to different application aspects, e.g., with ontology based specification in [Krieg-Brückner et al., 2004]. Similar to the principles of the route graph, conceptual models with different levels of information were proposed for various applications, e.g., [Zender et al., 2008, Martínez Mozos, 2010] developed and improved a topological information based multi-layered conceptual model corresponding to the spatial and functional properties of typical indoor environments to support a mobile robot's indoor navigation.
Furthermore, by considering the formal algebraic properties of qualitative spatial knowledge and its important role in spatially-related interaction, much research has also been carried out in qualitative spatial representation and reasoning. Based on the most prominent spatial calculi such as the cardinal direction calculus ([Frank, 1996]), double cross calculus ([Freksa, 1992]), region connection calculi ([Cohn et al., 1997]) and many others, general or domain specific QSR frameworks and models have been developed and proposed to support various spatially-related applications, e.g. [Wallgrün et al., 2007] proposed SparQ, a general toolbox for qualitative spatial reasoning in applications; [Bhatt et al., 2011] developed a declarative spatial reasoning framework and demonstrated its applicability for the domain of computer aided architecture design; many applications based on qualitative spatial knowledge have also been involved with human system interaction, such as in [Shi et al., 2006], the double cross calculus based spatial actions have been used as the fundamental unit in processing

natural language route instructions and interpreting them by fuzzy operations on a Voronoi graph to support human-robot collaborative spatial navigation; or in [Schultz et al., 2006] qualitative spatial reasoning has been integrated into query tools that are used by non-expert users in geographic information systems; or more recently, [Bellotto et al., 2013] proposed a qualitative trajectory calculus based approach to abstract and design robot behaviours for spatial interactions with a mobile robot.

1.2.3. Interaction in AAL

The mechanisms of typical interaction, either with single or multiple modalities, are usually only suitable for users with sufficient familiarity with information technology; while the potential user group in the Ambient Assisted Living environments consists mostly of elderly persons or persons with physical and mental impairments. Therefore, special focus has been given to research of interaction while taking AAL-centered characteristics into account from different perspectives.

Empirical studies have been conducted to collect objective and subjective data to motivate and support the development and improvement of interaction and interactive systems in AAL environments. For example, [Takahashi et al., 2003] reported a 'Wizard of OZ' (WOZ) experiment where elderly persons interacted with a home health care system and provided hints for natural language understanding for elderly persons; or in [Möller et al., 2008], dialogue corpora were obtained from interactions of older and younger users with a smart-home system, and the analysis results confirmed the significant difference of the two groups regarding either speaking style or vocabulary; or in [Ivanecky et al., 2011] empirical studies were also conducted on the usability of a mobile phone used by elderly or disabled people as the communication medium to control intelligent house environments and provided proof for the feasibility of the interactive system.

Combining empirical results and the AAL-centered characteristics, several efforts have been invested into developing and adapting different modalities to support interaction within AAL environments. For example in [Becker et al., 2009] a voice recognition system was developed within an assisted environment deployed with multiple sensors to build a health care monitoring system for elderly persons; or in [Goetze et al., 2010] acoustic user interfaces were developed especially for elderly persons in the context of AAL, and the implementation was demonstrated with a multi-media reminder and calendar system. As another important modality, intuitive gestures in the AAL context were investigated in [Nazemi et al., 2011] for identifying common interaction scenarios in an AAL environment with elderly persons; simple gesture-based interaction was also developed and integrated into a framework featuring three dimensional acceleration sensor information of WiiMote from Nintendo to be used in smart home environments ([Nesselrath et al., 2011]). Furthermore, new interfaces have also been developed to meet the special requirement of severely disabled persons, e.g. in [Mandel et al., 2009], a brain computer interface has been developed and used by disabled persons to steer an automated wheelchair.

Moreover, in order to generally improve the accessibility, flexibility and usability of interaction in AAL environments, considerable research effort has also been concentrated on developing multimodal interaction. Some focused on multimodal inputs, e.g., [Goetze et al., 2012] proposed a mobile communication and assistance system on a robot platform featuring acoustic, visual and haptic input modalities to be used by elderly persons in home-care envi-

ronments. Some focused on multimodal output, such as in [Boll et al., 2010], a multimodal reminder system using different acoustic, visual and tactile outputs as system responses was developed and used by elderly persons in their residential home. In addition, as the most basic aspect of multimodal interaction, modality fusion is performed in different ways, e.g., in the previous two examples, the multimodal fusion were implemented at the dialogue management level; while some others tried to perform the fusion at the grammar level, by integrating formal grammars and logical calculus as a multimodal language specification (e.g., [D'Ulizia et al., 2007]). Further research effort has also been made based on this principle of fusing multimodal events at the grammar level, e.g. [D'Andrea et al., 2009] proposed a multimodal pervasive framework for Ambient Assisted Living using multimodal grammar specification to support the interpretation of multimodal input, the management of the multimodal interaction and the generation of multimodal output.

1.3. Contributions of this Work

This dissertation investigates the design, development and implementation of **formal language based**, **spatially-related**, **multimodal** interaction in **AAL environments**. According to the research work addressed in the three research packages introduced above, the major contributions of the reported work are summarized as follows:

Formal Interaction. In general, this research package has been focusing on the modelling and management of interaction, in both theory and practice.

The first contribution of this package is the solution process centering on a formal language based dialogue modelling approach as well as the development of a formal language based computational framework for dialogue modelling and management called FormDia, the Formal Language Based Development Toolkit (see Section 2.1). Interaction processes, with either single or multiple modalities, can be specified using the formal language CSP ([Hoare, 1978, Roscoe et al., 1997]) as an abstract interaction model; then the CSP specification can be validated with the model checker FDR2 ([Roscoe, 1994, (Europe), 2010]) and verified with the simulator provided by the FormDia framework; and finally, the validated and verified model can be integrated into a practical interactive system to support formally tractable and extensible interaction management.

As the second contribution, by considering the limitations of conventional finite-state transition based dialogue modelling approach and the classic agent-based theory, i.e., the information state update based method, a unified dialogue modelling approach is developed (see Section 2.2). This approach benefits from the advantages of both classic models and can support an easily tractable, flexible and context-sensitive interaction in any complex interactive systems. With the FormDia framework, several unified dialogue models were accordingly developed, implemented and integrated into the interactive systems covered in the other two research packages of this thesis. Furthermore, the unified dialogue model implemented into a multimodal interactive system was especially evaluated through an empirical study. The effectiveness of the unified dialogue model was highlighted by the positive empirical results based on a standard statistical method.

Spatial Interaction. In general, this research package has been treating the following aspects in Human-Robot Interaction and Cognitive Science in depth: the management and formalization of, and reasoning with, spatially-related knowledge.
The most important contribution of this package is the development of a general four-level conceptual model based on qualitative spatial representation and reasoning (see Section 3.1). This model is called QSBM, the Qualitative Spatial Beliefs Model. It is used to support effective, efficient and user-friendly interaction between human operators and mobile robots while performing spatial navigation tasks. Specifically, the conceptual model can be used by mobile robot systems to represent spatial environments based on qualitative spatial knowledge; with the model based qualitative spatial reasoning, application-dependent low-level update rules can be implemented to manage the state of the situated environment; based on the low-level update rules, model based conceptual strategies can be developed for high-level spatially-related human-robot interaction; and finally, high development flexibility and extensibility are also ensured by the multi-level structure to support broader application possibilities.
As the next contribution, a DCC-based QSBM was developed by combining the conventional route graph ([Werner et al., 2000]) and the double cross calculus ([Freksa, 1992]) (see Section 3.2). As a result, this model benefits from the topological structure of a route graph for global navigation and the qualitative spatial DCC relations for intuitive communication with human operators. Accordingly, a set of low-level update rules were defined based on qualitative spatial representation and reasoning of DCC. These rules can refer to the most atomic route instructions one can use to instruct a mobile robot. Furthermore, with respect to the principle of general QSBM, a set of high-level conceptual strategies was developed and applied to manage the low-level update rules. Finally, these strategies are used to generate clarification dialogues for resolving different frequently occurring conceptual mode confusions caused by the disparity between the human's mental and the robots' internal representations of spatial environments.
To support the implementation of the QSBM, the low-level update rules as well as the high-level conceptual strategies, a computational framework called SimSpace was developed (see Section 3.3). On the one hand, SimSpace can be used as a stand-alone system for implementing, visualizing, simulating and testing QSBM-based instances of spatial environments and the QSBM-based functions; on the other hand, SimSpace is also well-encapsulated as a domain-dependent model-component, i.e., it can be directly integrated into an interactive system and used by a mobile robot to support spatially-related interaction with human operators.
Last but not least, empirical studies were conducted to evaluate an interactive system that implemented the QSBM-based models and functions (see Section 3.4). The evaluation was conducted especially for comparing the implemented set of high-level conceptual strategies. The positive results regarding effectiveness, efficiency and user satisfaction about the interactive system confirmed the important contributions of the QSBM-based model, the computational framework SimSpace and the conceptual strategies.

Multimodal Interaction in AAL. In general, this research package has been dealing with the following aspects in Multimodal Interaction and Ambient Assisted Living with

considerable effort: the development, and evaluation of, multimodal and AAL-centered interaction.

The first contribution of this package is the general support of design, development and implementation of multimodal interaction for elderly persons in AAL environments (see Section 4.1). As the general foundation of this contribution, two important aspects were highlighted: a) a list of elaborated design and development guidelines based on the consideration of the traditional design principles of conventional multimodal interactive systems, and the most common age-related decline of sensory, perceptual, motor and cognitive abilities of elderly persons; and b) a formal language supported unified dialogue model that combines a recursive transition network based generalized dialogue model and a classic agent based management method, which is used to support flexible and context-sensitive, yet formally tractable and extensible multimodal interaction for elderly persons. According to the two development foundations, MIGSEP, the Multimodal Interactive Guidance System for Elderly Persons was developed and implemented (see Section 4.2). The MIGSEP system runs on a portable touch-screen tablet PC and serves as the interactive assistant; it is intended to be used by an elderly or handicapped person seated in an electronic wheelchair that can navigate its user within complex spatial environments autonomously.

As the second contribution (see Section 4.3), via the cooperation with the department of medical psychology, medical sociology and neurology at the university medical center Göttingen, a series of empirical studies was conducted with groups of elderly persons. These studies systematically evaluated the minutely developed multimodal interactive system with respect to the touch-screen, the spoken language and the combination of both as input modalities, while enabling a continuously improved development process with respect to the subjective and objective data of each empirical study. Furthermore, a general model based evaluation framework was accordingly developed and proposed to analyze and compare the empirical multimodal data. The overall positive results showed high effectiveness of task performance, high efficiency of interaction and good user satisfaction with the interactive system. These findings also provided proof of the systematically developed and empirically improved design and development guidelines, foundations, interaction models and frameworks for supporting effective, efficiently and elderly-friendly multimodal interaction in AAL environments.

Chapter 2.

Formal Interaction

As the principle aspect of the dissertation, this research package investigated the design and development of general models and frameworks to support robust, formally tractable and manageable, flexible and context-sensitive interaction. On the one hand, these models and frameworks can be applied to any possible application domain involved with interaction or interactive systems; on the other hand, they can also be used to support the development of the interaction management component for the other two research packages in this dissertation. In this package, the major work effort has been concentrated on the two essential research aspects: a) a complete solution process featuring a formal language based dialogue modelling and management framework to enable the development of a formally tractable and manageable interaction modelling and management; and b) the development of a unified dialogue modelling approach that combined the generalized dialogue modelling and the classic agent-based information state update management theory to support an easily-tractable, flexible and context-sensitive interaction.
This chapter briefly introduces the contributed work as follows: the solution process with the formal language based dialogue modelling and management is presented in section 2.1, the development and implementation of the unified dialogue modelling and management approach is introduced in section 2.2, then the corresponding publications contributing to this research package are summarized in section 2.3, and finally the possible future work related to this package is given in section 2.4.

2.1. The Formal Language based Dialogue Modelling and Management

Correctness and robustness are two of the most important properties of interaction or interactive systems. However, to test whether an interactive system is correct or robust is usually a cumbersome and costly process. As introduced in subsection 1.2.1, interaction models can be represented as finite state transition networks. Meanwhile, formal languages can be used to specify finite state transition networks, and the formal language specification can be analyzed, tested and validated by theorem provers and model checkers (cf. [Roscoe et al., 1997, (Europe), 2010]). Therefore, an interaction modelling and management solution process featuring a formal language based development framework is developed and used to support the design, development and implementation of a formally tractable and manageable interaction and its integration into interactive systems.

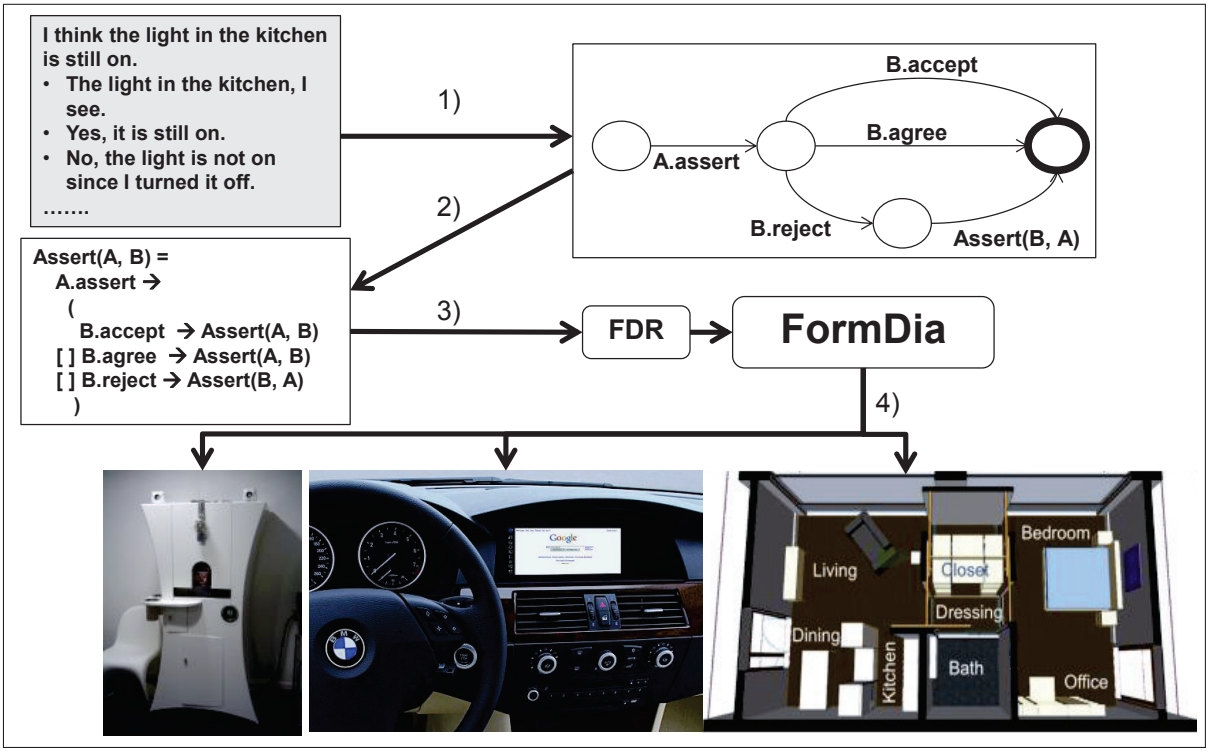


Figure 2.1.: The formal language based solution process.

2.1.1. The Formal Language based Interaction and Modelling Solution Process

Figure 2.1 illustrates the formal language based interaction modelling and management solution process, which consists of the following four important steps:

1) Semantic Interaction Modelling based on empirical data, interaction models can be constructed and illustrated as finite-state transition networks with straightforward interaction structures, which also ease the development process of semantic interaction models in a preliminary manner. The semantic models can be quite corpus dependent and contain details about the interaction context within the given corpus, or built at the illocutionary level without references to any direct surface indicators (cf. e.g. [Sitter and Stein, 1992]).

2) Formal Specification To bridge the gap between semantic models and machine readable codes, the formal language Communicating Sequential Process (abbr. CSP cf. [Hoare, 1978]) is used to specify the semantic model based finite state transition networks with abstract, yet highly readable and easily maintainable logic formalization (see the sample CSP specification of the simple semantic model in the step 2) of figure 2.1).

3) Testing and Validation CSP specifications can be loaded into the model checker FDR ([Roscoe, 1994]) for validating the functional concurrent properties, enabling the further development and improvement of the CSP specifications, and therefore increasing the tractability of the semantic interaction models.

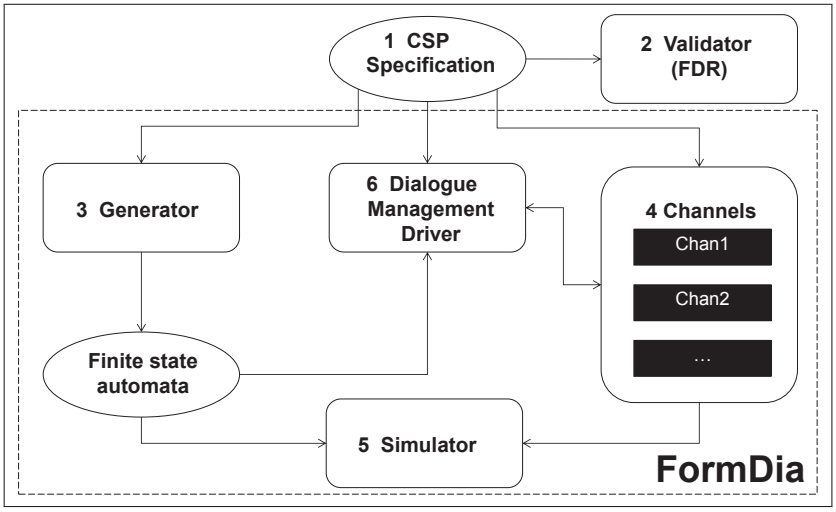


Figure 2.2.: The architecture of the FormDia framework.

4) Integration Finally, the CSP specifications can be imported into the formal dialogue development framework (abbr. FormDia) to support a further verification, simulation, development and implementation process as well as a direct integration into interactive systems such as the shown speech-enabled home device (e.g. [Heise.de, 2009]), a simple interaction assistant for an ambient assisted living environment (e.g. [Krieg-Brückner, 2013]), or to enable multimodal interaction in vehicles. The concrete details about FormDia is introduced in the next subsection.

2.1.2. The Formal Dialogue Development Framework

Based on the previous research work on the development and implementation of formal language based dialogue models (cf. [Shi et al., 2005, Shi and Bateman, 2005]), the formal dialogue development framework (abbr. FormDia) was further developed. Figure 2.2 illustrates the improved architecture of the FormDia framework. The current FormDia comprises six functional resources/components according to the development process of a formal language based dialogue model in a practical perspective, which includes its development, implementation and integration as an interaction management component into a practical interactive system. Specifically:

1. **CSP Specification** As introduced e.g., in figure 2.1, every dialogue model can be illustrated as a finite state transition network and accordingly, the structure of the finite state transition network can be specified as a CSP specification, i.e., a machine readable CSP program.

2. **Validator** the CSP specification can be validated by the model checking toolkit called Failures-Divergence Refinement (abbr. FDR cf. [(Europe), 2010]). This toolkit can be used to validate the functional properties of any CSP specification.

3. **Generator** according to the validated CSP specification, machine readable finite state automata can be generated by the Generator.

4. **Channels** based on the finite state automata, communication channels regarding all the generated finite states can be defined with domain specific information and handling mechanisms. These channels are only black boxes at the beginning, which will then be implemented with deterministic behaviour of concrete components with respect to their application contexts.

5. **Simulator** uses the generated finite state automata to simulate dialogue scenarios via an external graphical interface (uDrawGraph, cf. [BKB, 2005]), which can visualize the dialogue model as a finite state transition network based directed graph. With the corresponding communication channels, either black boxes or implemented ones, a set of utility functions are also provided by the Simulator to generate dialogue events and trigger the state transition for the advanced verification of the dialogue model within simulated dialogue scenarios.

6. **Dialogue Management Driver** after the validation and verification, the dialogue model based finite state automata and the communication channels are integrated into the dialogue management driver, which can then be directly used by a practical interactive dialogue system as the interaction management component.

The FormDia framework can be used as a general interaction modelling framework to develop and implement a formal language based dialogue model to enable formally tractable and extensible interaction. Furthermore, the framework can also be used to support the unified dialogue modelling and management approach, by implementing the Dialogue Management Driver and the Communication Channels with information state update based components (see section 2.2).

2.2. the Unified Dialogue Modelling and Management Approach

As a typical finite state transition network based approach, generalized dialogue models (cf. [Sitter and Stein, 1992]) were developed by structuring dialogues at the illocutionary level (cf. [Alston, 2000]) to enable surface-independent dialogue modelling. However, this modelling approach is criticized for lacking flexibility of handling dynamic information exchange. Meanwhile, the information state update based dialogue management theory was proposed by [Traum and Larsson, 2003] and provides a powerful mechanism to deal with dynamic information and therefore achieves a context sensitive dialogue management. Nevertheless, such models are usually very difficult to manage and extend ([Ross et al., 2005]). Thus, a unified dialogue modelling approach was developed by combining the generalized dialogue models with the information state updated based theories.
Figure 2.3 illustrates how a unified dialogue model is developed based on a generalized dialogue model with information state update rules. Specifically:

- Figure 2.3 a) shows a simple generalized dialogue model as a finite state based recursive transition network (abbr. RTN). It describes the dialogue situations where an agent A is making an assertion at the beginning, followed by the agent B's reaction with one of the three possible actions: accepting, agreeing on or rejecting A's assertion; if B

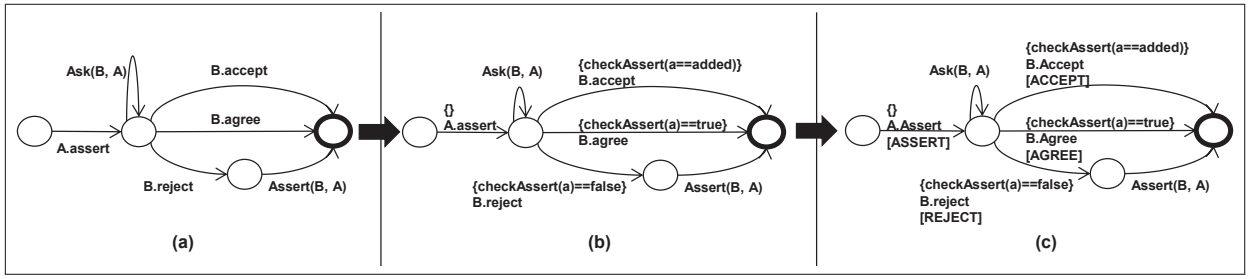


Figure 2.3.: The development process of a unified dialogue model.

rejects A's assertion, then B makes a follow-up assertion to A and triggers the recursive transition.

- The generalized dialogue model in figure 2.3 a) is a none deterministic model, i.e., no mechanism is defined about how B reacts to A's assertion. However, in order to build a feasible interaction model, deterministic behaviour should be assured for the interaction flow. Thus, conditional transitions are introduced to modify the original dialogue model into the one in figure 2.3 b), where checkAssert is a function to check whether an assertion holds within the knowledge base of B: if the assertion holds, B agrees with it; otherwise, B rejects it and initiates further discussion with a follow-up assertion; or if the assertion is not known by B, then B accepts it. As a result, the original dialogue model was modified as a conditional RTN with conditional transitions that can only be triggered if the relevant condition is fulfilled with respect to the concerned checking-function.

- Although the conditional RTN based generalized dialogue model specifies a deterministic illocutionary structure, it does not provide the mechanism to integrate discourse information, such as the assertion during the interaction. Thus, information state update based theory was accordingly applied by a) ignoring the typical element in the original information state update theory: i.e. the AGENDA for containing the next planned dialogue moves, since such information is already captured by the structure of the generalized dialogue model; b) complementing the illocutionary structure with information state based update rules, which are associated with the information state of discourse context and can update the information state respectively if necessary. As a result, a unified dialogue model is constructed as shown in figure 2.3 c), where four update rules: ASSERT, ACCEPT, AGREE, REJECT are added and used to access to the information state regarding context while performing updates accordingly. E.g. the update rule ACCEPT is used to add a new assertion into the knowledge base of B and considers this assertion as known by B from then on; or the update rule AGREE is used to insert the acknowledgement of the assertion into the topic under discussion.

In general, a unified dialogue model is developed as a recursive transition network with the following three essential features: a) it is built at the illocutionary level of interaction processes as a generalized dialogue model; b) its state transitions can only be triggered by fulfilled conditions concerning the information state; and c) a set of information state based update rules are defined and accordingly invoked during state transitions to update the information state if necessary. Therefore, a unified dialogue model benefits from both the

generalized dialogue model for a easily-tractable and manageable dialogue management and the information state update based theory for a flexible and context-sensitive interaction control.
With the introduced formal language based solution process and the FormDia framework in section 2.1, unified dialogue models can be developed and implemented with corresponding CSP specifications and domain-dependent channel drivers, and integrated into practical interactive systems for supporting unified dialogue model based interaction management. E.g., a unified dialogue model was used to incorporate spatial knowledge as information state context to support a spatially related interaction for human-robot collaborative navigation (see section 3.4); or another unified dialogue model was implemented into a multimodal interactive guidance system for elderly persons (see section 4.1.2) and evaluated with a series of empirical studies. Furthermore, an evaluation is conducted especially on the effectiveness of a unified dialogue model with a standard statistical method from the perspective of dialogue model level (see section 2.3);

2.3. Contribution of the Corresponding Publications

Major effort has been put into a general interaction modelling and management solution process with a formal language based dialogue modelling and management framework, as well as the development, implementation and empirical evaluation of a unified dialogue modelling approach to support a robust, flexible and context-sensitive, yet formally manageable and extensible interaction. Specifically,

- based on the previous work on using a formal method to support dialogue management ([Shi et al., 2005, Shi and Bateman, 2005]), a formal language based development toolkit for dialogue modelling was developed and improved, which enables an intuitive design of interaction models with a formal language, easy validation and verification for the formal language specified interaction models, and a straightforward integration into practical interactive systems ([7]).

- By combining the conventional recursive transition network based modelling and agent-based dialogue theory, a unified dialogue modelling and management approach was proposed. Then as a practical example, a unified dialogue model was implemented and integrated with the formal language based toolkit into a practical interactive system as the interaction model to support multimodal interaction ([7, 4, 5, 2]).

- using a standard statistical method, the kappa coefficient, the task success of the implemented unified dialogue model was evaluated through an empirical study in [11]. The positive results showed that the unified dialogue model is highly effective.

2.4. Possible Future Work

The reported work aimed to provide general methods, approaches and frameworks to support the development process of interaction management. Relating to the conducted research, further work effort can be concentrated upon the following aspects:

- Currently, the first step of the formal language based solution process, i.e., the semantic modelling of interaction, is usually performed in a hand-crafted way, where developers construct the semantic models based on the subjective evaluation of empirical data. In this situation, not only difficulties can arise with larger empirical corpora, but unforseen modelling errors are also likely to occur due to individual subjectiveness. Therefore, machine learning techniques can be applied to learning the semantic models out of empirical data, e.g. with the semantically annotated empirical corpora based on the work of [Shi et al., 2010].

- Although the unified dialogue model can support a flexible and context-sensitive interaction management with the integrated information state update theory, interaction with adaptive behaviours is gaining increasing interest in the recent years. Much research has been focusing on using reinforcement learning to optimize interaction behaviour either with collected empirical data or during the interaction with real users (cf. [Lecoeuche, 2001, Li et al., 2009, Pietquin et al., 2011]). Based on this research work, different reinforcement learning methods can also be implemented into the dialogue management driver of FormDia to support not only formally tractable and manageable, but also context-tailored adaptive interaction.

Chapter 3.

Spatial Interaction

This research package has focused essentially on the following two important issues in the scenarios of human-robot collaborative spatial navigation:

- To control the route, along which a mobile robot should go, human operators usually use natural language instructions that contain only qualitative spatial relations and conceptual landmarks (see e.g., [Werner et al., 1997, Hirtle, 2008]); while the mobile robot uses quantitative representation as the internal model about the spatial environment and it can usually accept route instructions consisting of only quantitative data;

- It is a rather complex process for human operators to provide a sequence of instructions to a mobile robot for route planning, since spatially-related communication problems could easily occur if a route direction is mistakenly given or spatial objects are incorrectly localized ([Reason, 1990, Bugmann et al., 2004, Marge and Rudnicky, 2010]).

Therefore, in order to a) bridge the communication gap between human operators and mobile robots, which is caused by the qualitative and quantitative disparity of their representation about space, and b) to support the collaborative negotiation of spatially-related communication problems in the sequence of route instructions, a qualitative spatial knowledge based four-level conceptual model: the Qualitative Spatial Beliefs Model (abbr. QSBM) was developed and used as the foundation of this research package for supporting intuitive human-robot spatially-related interaction.
This chapter briefly presents the major focus of this research package as follows: the general Qualitative Spatial Beliefs Model is introduced in section 3.1, followed by a qualitative spatial calculus dependent instance of QSBM in section 3.2, then a computational framework that implements the QSBM model and the model based functions is presented in section 3.3; empirical studies were conducted regarding the resolving of the spatially-related communication problems using QSBM and reported in section 3.4; finally, the contribution of the corresponding publications is summarized in section 3.5 and the possible future work is given in section 3.6.

3.1. QSBM: A general Qualitative Spatial Beliefs Model

From the perspective of human operators, spatial environments are not represented with quantitative data as a mobile robot does, but with conceptual objects or places and their

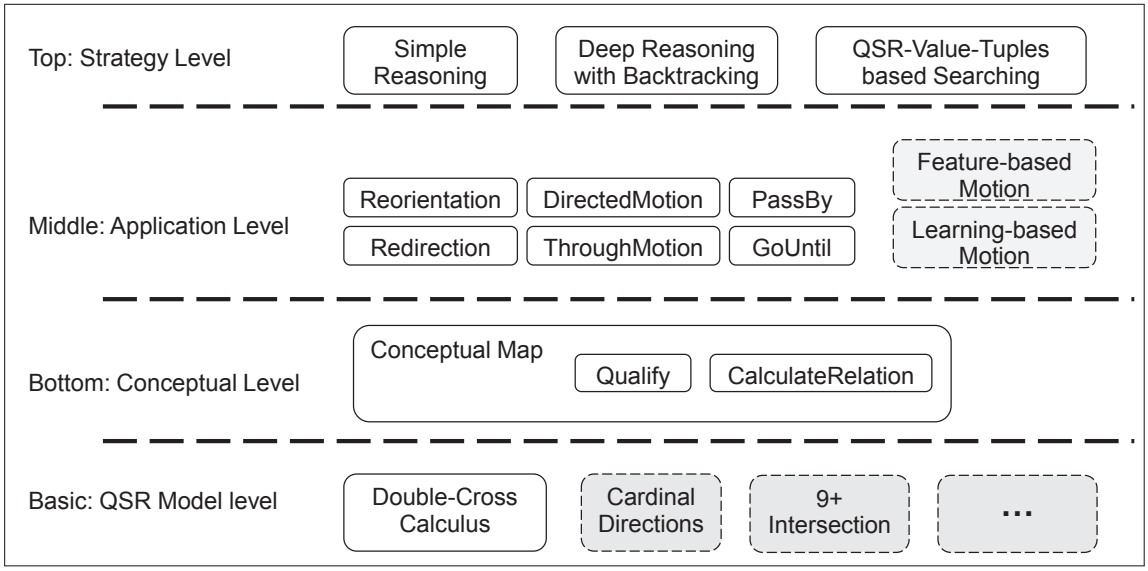


Figure 3.1.: The General Qualitative Spatial Beliefs Model.

qualitative spatial relations. For human operators to communicate with mobile robots for spatial navigation tasks, an intermediate knowledge representation is needed and accordingly, Qualitative Spatial Beliefs Model (QSBM), a qualitative spatial knowledge based four-level conceptual model, was developed to model a mobile robot's beliefs to support human-robot collaborative navigation. The general QSBM is illustrated in Figure 3.1 and introduced as follows:

- The basic level **QSR Model level** refers to the most basic theoretical foundation of a QSBM: the qualitative spatial calculi to meet the requirement of different application scenarios, such as Double-Cross Calculus ([Freksa, 1992]), Cardinal Directions ([Frank, 1996]), 9+ Intersection ([Kurata, 2008]), etc.

- The bottom level **Conceptual Level** holds the fundamental conceptual model of the QSBM: a conceptual map with only conceptual objects and qualitative spatial relations regarding a chosen qualitative spatial calculus. It facilitates the most basic calculating and reasoning operations related to the connection between the chosen calculus and the spatial environment. It is used as a black box holding a conceptual qualitative spatial knowledge based representation of a spatial environment. It provides two basic functions: *Qualify* for qualifying quantitative data into calculus-based qualitative relations, and *CalculateRelation* for calculating qualitative spatial relations between objects using calculus-based qualitative spatial reasoning.

- The middle level **Application Level** consists of a set of application-dependent update rules corresponding to the most atomic route instructions one can use to instruct a mobile robot in human-robot collaborative spatial navigation. For instance, the update rule *Reorientation* refers to the instruction "turn left", *Redirection* interprets "take the next junction on the right", *Feature-based Motion* concerns instructions with features of objects or landmarks, such as "go around the big room" ([12]), and *Learning-based Motion* represents those instructions enabling the robot to update its conceptual

knowledge by acquiring new landmarks from given instructions, such as "the third office is the directory's office, pass by it", etc. According to the formal definition, each update rule is used to update the conceptual map on the conceptual level, based on a chosen calculus and the related qualitative spatial reasoning on the QSR Model level.

- The top level **Strategy Level** includes a set of high-level conceptual strategies for interpreting a sequence of route instructions and if possible, providing qualitative spatial knowledge based information to resolve the spatially-related communication problems during the human-robot collaborative spatial navigation. In practice, each conceptual strategy defines its own mechanism for appropriately choosing and applying atomic update rules on the application level to update the conceptual map on the conceptual level.

In general, QSBM a) provides a conceptual model with qualitative spatial knowledge to represent spatial environments; b) applies qualitative spatial reasoning to support application-dependent low-level update rules to update the conceptual representation; and c) offers high-level conceptual strategies to manage the atomic update rules to support high-level spatially-related human-robot collaborative navigation. Benefiting from the flexibility and extensibility of the multi-level structure, different qualitative spatial calculi can be used on the QSR model level to support various application scenarios, application-dependent atomic actions can be easily defined and extended on the application level, or different high-level conceptual strategies can also be developed with respect to different ways of applying update rules for resolving communication problems, while each of these changes/extensions requires only limited adaptation on the other levels in QSBM.

3.2. A DCC-based QSBM

According to the requirement of the focused scenario of this research package, double-cross calculus (DCC) is used on the basic QSR model level. Accordingly, a DCC-based Qualitative Spatial Beliefs Model is developed and a brief introduction to the other levels of the DCC-based QSBM is given as follows:

3.2.1. The Conceptual Level

On the one hand, as a common knowledge base of a mobile robot involved in spatial navigation, the Route Graph, was proposed in [Werner et al., 2000], which models the conceptual topological knowledge on the cognitive level in navigation space from human's perspective. Route graphs can be used as metrical maps in global navigation for mobile robots and ease the interaction with human operators to a certain extent. However, lacking qualitative spatial relations between objects, conventional route graphs are not suitable for supporting natural language based human-robot collaborative navigation. On the other hand, Double-Cross calculus (DCC) was proposed in [Freksa, 1992] for qualitative spatial representation and reasoning using the conceptual orientation grids, where a directed segment divides the 2-dimensional space into disjoint grids and can define 15 meaningful qualitative spatial relations. DCC model can therefore describe the relative relations between objects in the local navigation map from an egocentric perspective. However, the conventional DCC model does

not consider the topological relations within global navigation maps.

To benefit from the two well-accepted conceptual models, the conceptual route graph (abbr. CRG) was developed by combining the topological structure of conventional route graph and the conceptual orientation grids of Double-Cross Calculus. Instead of quantitative information, DCC qualitative spatial relations are used to describe all the relative relations between route graph nodes and route graph segments. Formally, a CRG is defined as a tuple of four elements, (M, P, V, R), where

- M is a set of conceptual landmarks in a spatial environment, each of which is located at a place in P.
- P is a set of topological places on the conceptual level of a spatial environment.
- V is a set of vectors from a source place to a target place, both of which belong to P
- R is a set of DCC based qualitative spatial relation-pairs, describing the qualitative spatial relations between each place and related vectors that define the orientation grids of DCC.

As a simple example, a CRG can be represented as the following specification:

```
crg = ( M = {kitchen : p1, printer: p2},
        P = {p1, p2, A, B},
        V = {AB, BA},
        R = {<AB, RightFront, p1>, <AB, LeftBack, p2>} )
```

This specification indicates that, this is a spatial environment containing two landmarks: kitchen, located at p1, and printer, located at p2, and two vectors AB and BA, with the relation-pairs showing that, kitchen is at the right-front of AB and printer is at the left-back of AB.
The model of Conceptual Route Graph provides a semantic framework for supporting human-robot collaborative navigation with the intuitive interpretation of human route instructions as well as the straightforward presentation of internal feedback from a mobile robot with the DCC-based qualitative spatial representation and reasoning, meanwhile it can also be used as a direct interface with the low-level mobile robot system for performing navigation tasks via the topological structure of the conventional route graph.
With the conceptual route graph as the conceptual map on the conceptual level of the DCC-based QSBM, the state of a DCC-based QSBM can be specified as a tuple of two elements: (crg, pos), where "crg" represents the conceptual route graph, and "pos" represents a vector of the current position and orientation of a mobile robot in the given conceptual route graph.

3.2.2. The Application Level

In order to support human-robot spatially-related interaction, natural language based route instructions from a human operator should be interpreted to update the state of a mobile

robot's QSBM instance, so that possible feedback regarding the interpretation can be transferred back to the human operator. According to the empirical studies on human-robot collaborative navigation (cf. [Bugmann et al., 2004, Roger et al., 2007, Shi and Tenbrink, 2009]) and the previous research effort related to natural language, cognitive models and route instructions (cf. [Denis, 1997, Tversky and Lee, 1998, Lauria et al., 2002]), a set of update rules regarding the most atomic route instructions one can use to instruct a mobile robot were developed.
Formally, each update rule is defined with the following three elements:

- *RULE* refers to the name that identifies an atomic type of route instructions.
- *PRE* is a set of preconditions, under which this update rule can be applied.
- *EFF* describes how the state of the QSBM is updated after applying this update rule.

For brevity two update rules are presented as examples as follows, while the other update rules can be found in the contributing publications in section 3.5.

Reorientation refers to the simplest route instructions, which is used to change the orientation of a robot regarding its current position. "Turn left", "Turn right" and "Turn around" are the typical expressions of such instructions. The precondition for Reorientation is whether the robot can find a CRG place in the current state of QSBM with the following two conditions: 1. it is connected with the current position, and 2. it has the desired spatial relation with the current position; the effect is that the robot faces that found CRG place after the reorientation. Formally it is described as:

```
RULE:  Reorientation
 PRE:  pos = ab,
       ∃ ac ∈ V . <ab, dir, c>
 EFF:  pos = ac
```

Concretely, this rule indicates that the robot is currently at the place a and facing the place b (ab is a CRG vector), if there exists a CRG vector ac with a targeting place c, such that the spatial relation of c with respect to the vector ab (i.e. the current position) is the desired direction dir to turn, i.e., <ab, dir, c>, then the current position will be updated as ac after applying this update rule.

Passing Motion relates to the route instructions containing an external landmark to be passed by, e.g. "pass the kitchen" or "pass the printer on the right" with directional information. For these route instructions, the robot should first identify the landmark and then check whether the landmark can be passed by along the current directed path. Furthermore, the desired passing direction should be considered as well, if the direction for passing the landmark is given. Accordingly the update rule PassLeft for passing a landmark on the left is specified as:

```
RULE:  PassingLeft
 PRE:  pos = ab,
```

```
        ∃ cd ∈ V . (landmark : l)
                 ∧ <ab, LeftFront, l> ∧ <cd, LeftBack, l>
                 ∧ <ab, Front, c> ∧ <ab, Front, d>
EFF:  pos = cd
```

This rule tries to find if there is the desired landmark and a vector cd, such that the landmark is located at the place l, which is on the left front of the robot regarding the current position ab, and left behind the robot with the updated route segment cd after executing the update rule.

3.2.3. The Strategy Level

With the update rules defined on the application level, single route instructions can be interpreted. However, in human robot collaborative navigation, instead of giving one single instruction, human operators usually give a sequence of route instructions to the mobile robot. In this case, before human operators can organize the appropriate terms for giving the instructions, they first need to correctly locate the robot's current position and the desired goal location, and then take the imagined journey in mind to go along the expected route while encountering possible mental rotation during the travelling. In this complicated process, a wrongly located place or turning can happen quite often ([Shi and Krieg-Brückner, 2008, Shi et al., 2006]). These errors can cause the failure of the interpretation of the following route instructions and consequently lead to so-called conceptual mode confusion situations, where the mobile robot goes along an undesired path or even simply cannot execute the desired instructions. In order to cope with these problems, a set of high-level conceptual strategies was developed on the strategy level, which can choose and apply the low-level update rules on the application level according to different principles for resolving conceptual mode confusion.

Deep Reasoning can deal with one of the most typical conceptual mode confusions called *spatial relation or orientation mismatches*. This type of conceptual mode confusion occurs, if a spatial object is incorrectly located in the operator's mental representation, such as "pass the kitchen on the left", where the kitchen is currently located on the right; or "take the second junction on the left", where the second junction is only leading to the right.

In this situation, the strategy of deep reasoning finds the suitable low-level update rules, then checks the preconditions of chosen update rules with the currently observed state of the QSBM using qualitative spatial reasoning. If an unsatisfiable precondition is identified by an update rule, this situation can be presented back to the human operator appropriately, or furthermore, if possible, by checking the update rule corresponding to the route instruction leading to the unsatisfiable situation, a corrected spatial relation can be inferred to build a possible suggestion.

Therefore, this strategy tries to resolve the problematic situation by either giving a reason regarding the current spatial situation to support the human operator to reorganize the route instructions, such as "you cannot pass the kitchen on the left, because it is now behind you", or it can make a suitable suggestion if one exists, such as "you cannot take a right turn here, but maybe you mean to take a left turn?"

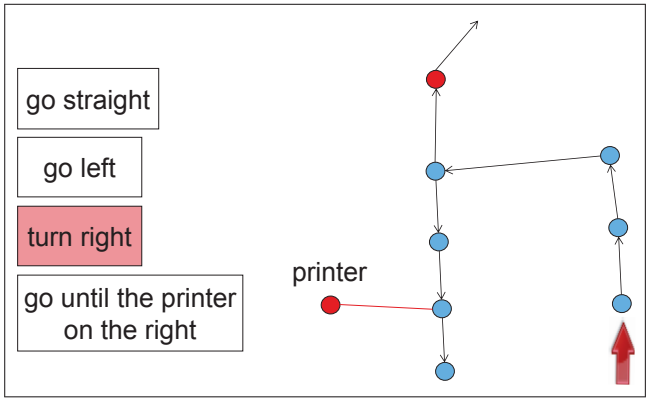


Figure 3.2.: A sequence of route instructions with a wrong instruction in the middle.

Deep Reasoning with Backtracking can not only handle the situations covered by deep reasoning, but also cope with situations where the failure of the interpretation of an instruction is caused by a previously wrong instruction, e.g. see the situation in figure 3.2. The robot is located at the thick red arrow and the instructions are: "go straight ahead, then go left, and then turn right, and go until the printer on the right." The check fails on interpreting the fourth instruction "go until printer on the right", because there is no kitchen ahead after taking a right turn as the previous instruction. However, by taking one step backwards, if the third instruction is changed from turning right to left, then the last instruction can also be interpreted accordingly.
Thus, this strategy interprets the route instructions as the deep reasoning does by checking every precondition of the chosen update rules. Yet after applying each update rule, the state of the updated QSBM is also saved in an interpretation history. Once one instruction cannot be interpreted, the previous state of the QSBM can be reloaded to replace the current state, then a possible suggestion can be made based on the previous instruction if possible, such as "turn left" instead of "turn right" in the example in figure 3.2. As a result, the interpretation of the remaining route instructions can be resumed based on the suggested route instruction, and if possible, instead of giving a reason or a suggestion regarding a certain problematic instruction, the deep reasoning with backtracking can manage to locate the previously wrong instruction and find a successful interpretation of the entire sequence of route instructions if such exists.

QSR-Value Tuples based Searching was developed to cope with a new type of conceptual mode confusion regarding wrongly located starting or turning position (called "conceptual turning location mismatch"). Figure 3.3a illustrates an example of this conceptual mode confusion, where the robot is located at the thick red arrow and the instructions are "go straight, then left, then go until the printer on the right". From the perspective of a human operator, the printer is located directly on the right side after taking a left turn, and therefore the operator simply ignores a turning point which is not in his or her mental representation. However, after taking a right turn, the last instruction "go until the printer on the right" cannot be interpreted, because there is no continuing possibility in the current state of the QSBM.
These problems cannot be solved by the other strategies, because they can only provide suggestions if there exists a wrong route instruction, while in this situation one

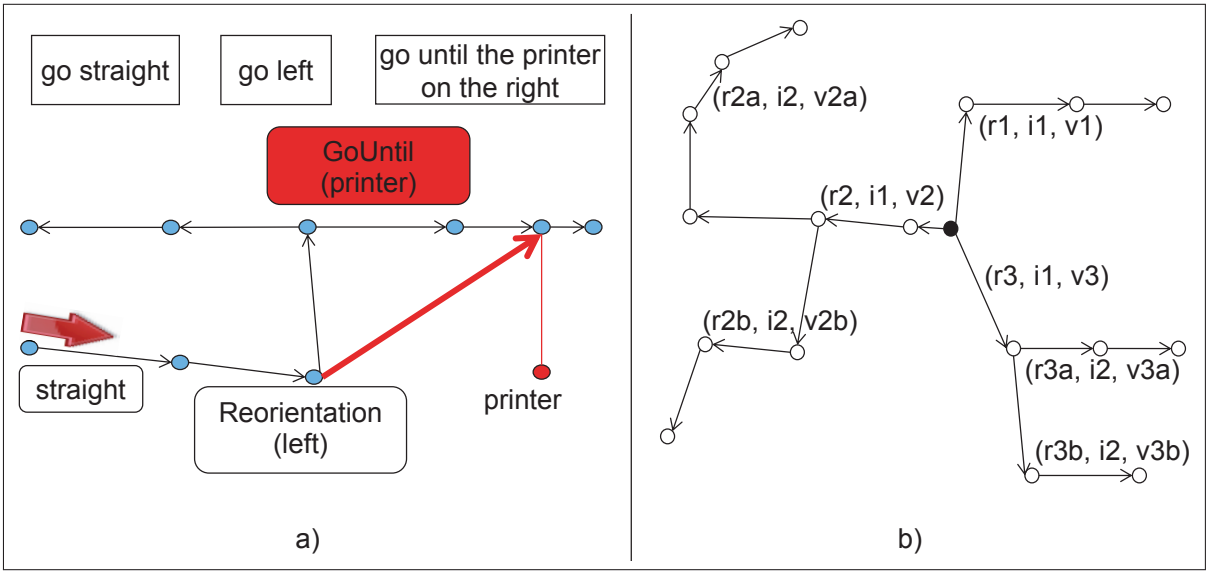


Figure 3.3.: a) An example of conceptual turning location mismatch; b) An abstract view of the QSR-Value Tuples based Searching.

route instruction (e.g. "turn right" as a third instruction) is missing. Thus, the strategy "QSR-Value Tuples based Searching" defines a QSR-weighted value tuple for each outgoing direction of each turning node in a conceptual route graph as:

(ROUTE, INSTRUCTIONS, QSR-V)

Here ROUTE represents the currently chosen route, INSTRUCTIONS includes all the interpreted instructions along this route, and QSR-V is the cumulative value calculated by

$$\text{QSR-V} = \sum_{i=0}^{i} MR_i * SR_i$$

where MR_i is the matching rate by comparing the taken qualitative spatial direction with the current route direction while interpreting the i-th route instruction, and SR_i is the success rate of interpreting the route instruction at that point.
With the definition of the QSR-weighted value tuple, finding an appropriate interpretation (namely a route) to correspond to a sequence of natural language route instructions is illustrated in in figure 3.3b as:

- An empty set of QSR-weighted value tuples was initialized at the current robot position (the black point in the middle of the network in figure 3.3b).
- This value-tuple-set is automatically updated by the QSBM manager in e.g. the 3 directions with $(r1, i_1, v1), (r2, i_1, v2), (r3, i_1, v3)$ in figure 3.3b, where (rx, i_i, vx) indicates the tuple of the covered route rx, the interpreted instructions i_i and the currently calculated QSR-weighted value vx).
- Searching agents of the QSBM manager are then travelling along all paths according to the branching of the current point on the current QSBM. The value-

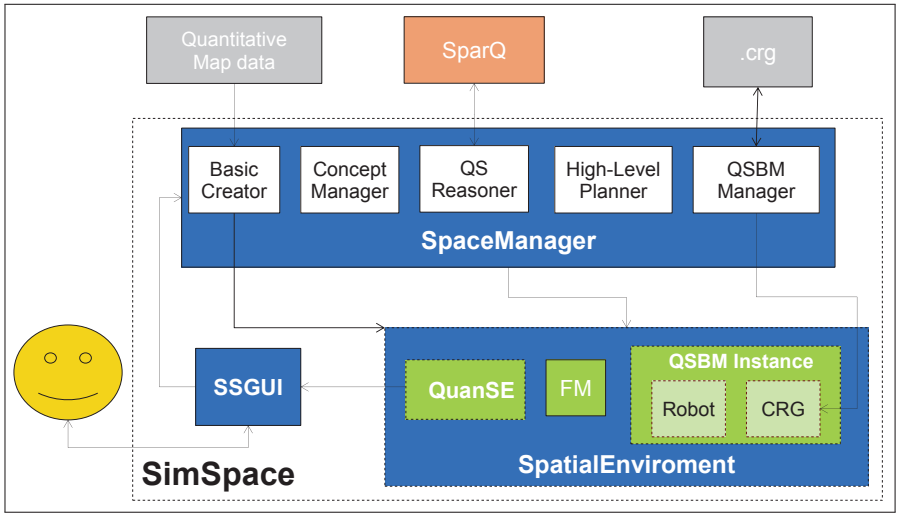


Figure 3.4.: The Architecture of SimSpace.

tuple-set gets updated and expanded by the QSBM based update rules when new branches are encountered or new instructions are interpreted.

- Finally, a full set of value-tuples is generated. The value tuple with the highest QSR-weighted value is either the best possible solution for interpreting the given route instructions or contains the most relevant information for possible suggestion/correction to resolve the conceptual mode confusions.

This strategy resolves conceptual mode confusion from a different perspective of a mapping problem in a directed graph with QSR weighted values compared to the other strategies; meanwhile, it also preserves the functionality of deep reasoning with the QSBM update rules and the QSBM instance on each outgoing path from each turning location. Therefore, it can be viewed as a searching algorithm with multiple deep reasoning agents to support the interpretation of more route instructions and clarifying more conceptual mode confusions.

3.3. SimSpace: A Computational Framework to Support QSBM

Based on the formal definition of QSBM, QSBM-based update rules and conceptual strategies, a conceptual model based computational framework SimSpace was developed for supporting the implementation, simulation and evaluation of QSBM.
According to the Model-View-Controller architecture (originally from [Burbeck, 1987]), the general architecture of SimSpace consists of a Model component SpatialEnvioronment, an optional View Component SSGUI and a Controller SpaceManager, with some external resources (see figure 3.4):

SpatialEnvironment keeps the current state of a QSBM instance, including the CRG instance and the hypothesis of a mobile robot, as well as the optional quantitative spatial

environment (QuanSE) with quantitative information and the optional feature map (FM) component with the conceptual information.

SSGUI is an optional interface component that is only used if SimSpace runs as a stand-alone application. It visualizes the spatial environment with quantitative and conceptual descriptions, interacts with a human user via natural language route instructions, and communicates with the SpaceManager by sending route instructions to be interpreted as well as receiving returned system feedback.

SpaceManager is the central processing component of SimSpace with five important functional components: a) BasicCreator for creating a QSBM instance with a given predefined specification of a certain spatial environment; b) ConceptManager for managing an ontology database of the conceptual knowledge for interpreting the conceptual terms in human-robot interaction; c) QS Reasoning for connecting the general framework SparQ ([Dylla et al., 2006, Wallgrün et al., 2007]) to implement the qualitative spatial representation and reasoning based functions on the QSR model level; d) QSBM Manager for defining the low-level update rules on the application level to interpret individual route instructions, manipulating and updating QSBM instances, as well as saving and loading QSBM instances into and out of an XML-based specification; e) High-Level Planner for implementing the high-level conceptual strategies on the strategy level to resolve conceptual mode confusions.

Generally, SimSpace provides two important functionalities: a) it can run as a stand-alone evaluation platform for visualizing spatial environments, generating corresponding QSBM instances and testing the interpretation of route instructions, and b) it is also a well encapsulated module, which can be integrated into an interactive mobile robot system for supporting the human-robot collaborative navigation with the qualitative spatial representation and reasoning on a human-friendly level while facilitating direct communication with a mobile robot via the inherited features from route graphs.

3.4. Empirical Studies of QSBM based Spatial Interaction

In order to evaluate the QSBM model regarding different conceptual strategies for supporting human-robot collaborative spatial navigation, QSBM, the QSBM update rules and conceptual strategies have been integrated using SimSpace within a practical interactive system and several empirical studies were accordingly conducted.
For all the studies, the interactive system was similarly set up as a networked software system consisting of two laptops: one laptop, called the system laptop, holds the actual interactive system with the graphical user interface with either hidden simulated maps or displayed maps of a real spatial environment, interaction manager based on a unified dialogue model (see section 2.2), speech synthesizer and the spatial knowledge processing component SimSpace that implemented the QSBM model; the other laptop, called the speech recognizer laptop, runs a graphical interface that was only operated by a human investigator and used to transfer the natural language instructions to the system laptop via wireless network. As a result, the whole system was simulated as if each participant was giving instructions to the system using spoken natural language directly.

The tasks for all empirical studies were also similar. Given several starting positions, the participants were asked to tell an avatar of a mobile robot to go to a certain goal location with sequences of route instructions, while interacting with the system about possible feedback concerning the frequently occurring conceptual mode confusions.
The positive empirical results from systematic evaluation have confirmed the effort on using the qualitative spatial knowledge based QSBM model to support effective, efficient and user-friendly human-robot collaborative navigation. Several comparisons between the high-level conceptual strategies were also conducted and discussed, which showed that the conceptual mode confusion can be resolved to a great extent with the conceptual strategies on the one hand; and on the other hand, a continuously improved development process of QSBM and QSBM-based functions is also ensured based on the empirical findings. For brevity the concrete details about the related studies can be found in the contributing publications (see section 3.5).

3.5. Contribution of the Corresponding Publications

As the central point of this package, major work efforts have been dedicated to developing and implementing a qualitative spatial knowledge based conceptual model to support the interaction process between human operators and mobile robots while resolving conceptual mode confusion in spatial navigation. Specifically,

- based on the previous work on conceptual modelling from the perspective of human operators ([Krieg-Brückner and Shi, 2006, Shi and Krieg-Brückner, 2008]), the hybrid qualitative spatial model called Conceptual Route Graph was implemented and integrated into a computational framework SimSpace to support the interpretation of natural language human route instructions with qualitative spatial reasoning [16].

- Based on the conceptual route graph, a qualitative spatial beliefs model was developed and reported in [14], which can be used as a general model for representing and reasoning with a mobile robot's internal knowledge about the spatial environment.

- With the qualitative spatial beliefs model, a set of high-level reasoning strategies were developed and implemented to enable mobile robots to validate natural language route instructions and generate corresponding clarification dialogues, and therefore assist human operators and mobile robots to identify, exemplify and reduce to a great extent the conceptual mode confusion in human-robot interaction ([13]).

- During the further development of the qualitative spatial beliefs model and the integration of the improved model-based computational framework SimSpace into a practical interactive system for a mobile robot, an additional type of conceptual mode confusion was identified and accordingly, a new high-level strategy was developed and presented in [3].

- Currently, the qualitative spatial beliefs model has been improved and generalized to a qualitative spatial knowledge based four-level conceptual model to enable broader application with e.g., other qualitative spatial calculi, further application-dependent actions or high-level conceptual strategies, etc. Together with the conceptual model

and the model-based framework, a practical interactive system using real environment maps has been implemented and evaluated in an empirical study. The study was concerned with different conceptual reasoning strategies for resolving conceptual mode confusion in the scenarios of human-robot collaborative spatial navigation ([1]).

Besides the major focus, the author also conducted further research work related to the area of interaction within spatially-related applications. E.g., [12] described the work on supporting inferences in wayfinding tasks in a multilevel building by selectively adding structural information. The empirical results provided interesting suggestions on developing an interactive wayfinding assistance system. And [9] studied the indoor route instructions with elaborate descriptive information based on an empirical evaluation of a natural language route direction system. The results highlighted the importance of a more flexible combination of prescriptive and descriptive information about spatial environments.

3.6. Possible Future Work

The reported work in this research package served as a continuing step towards building effective, efficient, user-friendly models and frameworks for spatially-related applications. Further work effort can be investigated from the current point of view, for example:

- Other qualitative spatial calculi can also be used and specified to build conceptual models on the QSR model level of QSBM to support different applications, such as using Cardinal Direction to support observation-centered scenarios (cf. [Wang et al., 2013]).

- [Marge and Rudnicky, 2010] stated that, quantitative data based expressions should also be considered in the understanding of spatial language. Therefore, benefiting from the quantitative information that is already contained in a Conceptual Route Graph, quantitative data based route instructions can be supported by accordingly adding low-level update rules into the application level, without affecting the other levels of QSBM.

- Learning-based route instructions on the application level can be investigated to support human-robot navigation within partially known or unknown environments, such as "the kitchen is the first room after taking a left turn, pass by the kitchen, then take a right.", which involves the combination of knowledge acquisition and real-time application.

- The strategy of QSR-weighted value based searching can be improved and optimized in some ways, e.g., a) some conceptual mode confusion is still only covered by the strategy of deep reasoning with backtracking. Therefore, an optimized version based on the two currently best conceptual strategies can also be implemented to get better coverage of the conceptual mode confusion; or b) the calculation of QSR-weighted values is currently based on predefined constants. By applying a reinforcement learning model based on continuously updated QSR-weighted values, this strategy will be able to find either user-adaptive or environment-adaptive interpretations of route instructions.

Chapter 4.

Multimodal Interaction in AAL

This research package has been concentrating on developing an effective, efficient and elderly-friendly multimodal interaction in ambient assisted living environments regarding the following two important aspects: a) the general support for the design and development of multimodal interaction for elderly persons as well as the implementation of a practical multimodal interactive guidance system to be used by elderly persons in autonomous navigation within complex buildings; and b) the development of a standard model based general evaluation framework concerning the effectiveness, efficiency and user satisfaction of multimodal interaction and its application in empirical evaluation of the implemented multimodal interactive guidance system for elderly persons.
This chapter gives a brief introduction to the major focus of this research package as follows: the two fundamental aspects for designing and developing multimodal interaction for elderly persons are presented in section 4.1; then according to the design and development foundation, MIGSEP, a multimodal interactive guidance system for elderly persons, was implemented and is described in section 4.2; a general evaluation framework for multimodal interaction and the application of the framework in empirical evaluation of the MIGSEP system while focusing on comparing the multiple input modalities are reported in section 4.3; finally, the contribution of the corresponding publications is summarized in section 4.4 and an outlook to the possible future work is given in section 4.5.

4.1. Foundation of Design and Development of Multimodal Interaction for Elderly Persons

The design and development foundation of this package consists of two aspects: a) a set of general guidelines for supporting the design and development of multimodal interaction for elderly-persons by taking the ageing-centered characteristics into account; and b) a unified dialogue model based on the unified dialogue modelling approach and formal method based framework (see chapter 2) to support the modelling and management of multimodal interaction.

4.1.1. Design Guidelines of Multimodal Interaction for Elderly Persons

A uniform decline in sensory, perceptual, motor and cognitive capabilities can by no means be assumed with increasing age, especially in the seven most common human abilities as shown in figure 4.1: *Visual Perception* worsens for most people while ageing, physically the

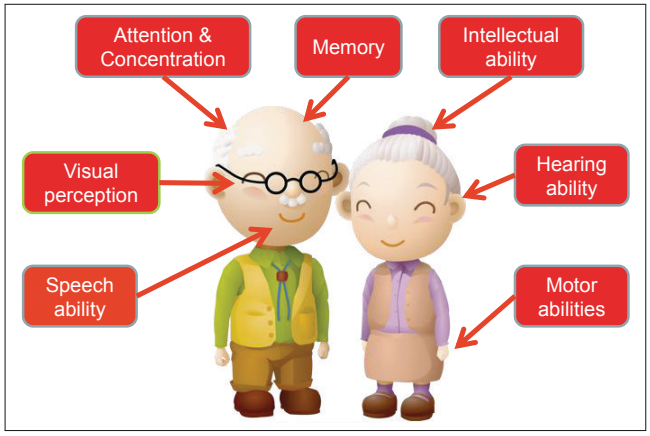


Figure 4.1.: The seven most common human abilities [10].

size of the visual field is decreasing and peripheral vision can be lost. It is more difficult to focus on objects up close and to see fine details, including identifying rich colors and complex shapes within images; rapidly moving objects are either causing too much distraction, or become less noticeable ([Fozard and Gordon-Salant, 2001]). *Speech Ability* declines while ageing in the way of being less efficient for language production in pronouncing complex words or longer sentences ([Mackay and James, 2004]), probably due to reduced physical functionality for controlling tongue or lips ([Burke and Mackay, 1997]); [Möller et al., 2008] also confirmed that elderly-focused adaptation of spoken language interface can improve the interaction quality to a satisfactory level. *Attention and Concentration* drop while ageing: elderly persons either become more easily distractible by details and noise, or find other things harder to notice when concentrating on one thing ([Kotary and Hoyer, 1995]); great difficulty has been shown with situations with divided attention ([McDowd and Craik, 1988]). *Memory Functions* decline with respect to different memory types. Short-term memory holds fewer items with age and working memory becomes less efficient ([Stoltzfus et al., 1996]), while semantic information however, is normally preserved in long-term memory ([Craik and Jennings, 1992]). *Intellectual Ability* does not decline much during the normal ageing, yet [Hawthorn, 2007] believed that because of crystallized intelligence elderly persons can perform better in a more stable well-known environment. *Hearing Ability* declines to 75% between the age of 75 and 79 ([Kline and Scialfa, 1997]). High-pitched sounds are becoming hard to perceive; complex sentences are difficult to follow ([Schieber, 1992]). *Motor Abilities* decline generally caused by loss of physical activities while ageing. It is more difficult to perform fine motor activities, such as grabbing small or irregular targets [Charness and Boot, 2009]; conventional input devices such as a computer mouse are less preferred by elderly persons since sufficiently good hand-eye coordination is required ([Smith et al., 1999]).
The above empirical findings and much more other research work on relating the effects of ageing with computer based systems (cf., e.g., [Ziefle and Bay, 2005], [Fisk et al., 2009], [Leung et al., 2012]) have clearly shown that it is necessary to consider ageing-centered characteristics while developing interfaces or interactive systems for elderly persons. Therefore, according to the common design principles for conventional interactive systems and the ageing-related characteristics regarding the seven most common human abilities, a set of guidelines for designing and developing multimodal interactive system for elderly persons

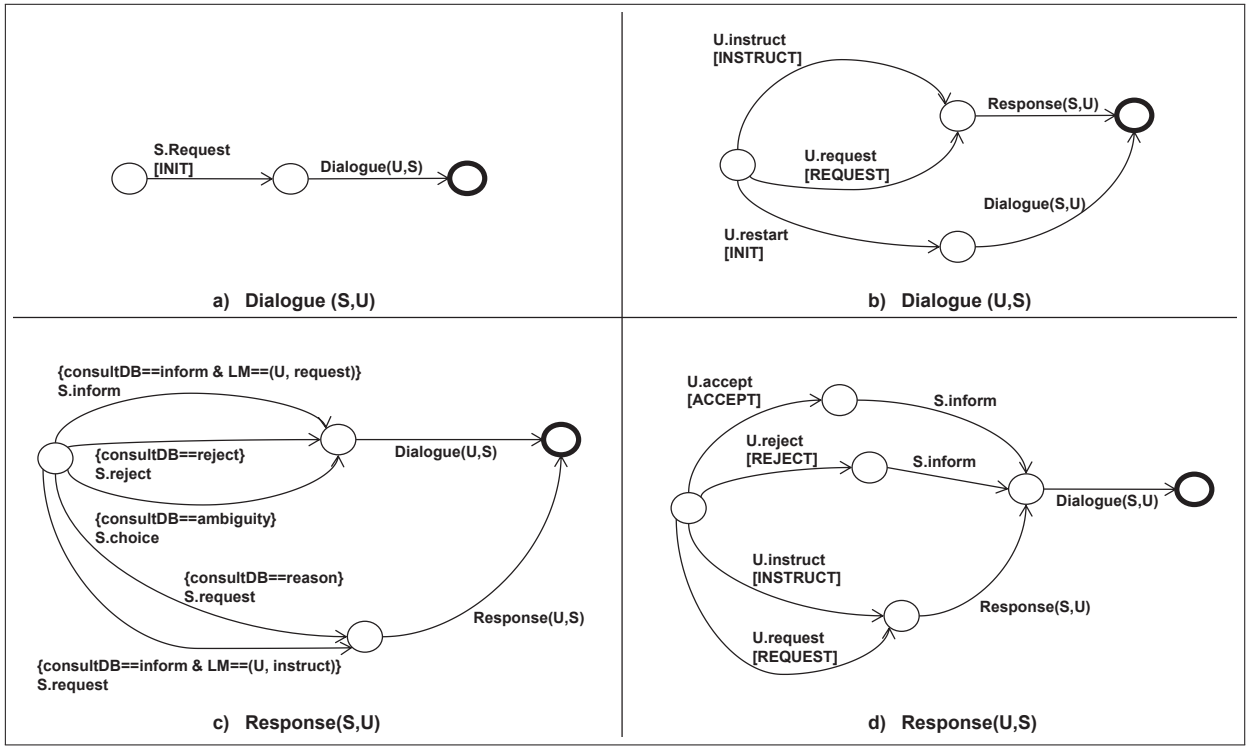


Figure 4.2.: The unified dialogue model comprising 4 sub-models: a) the initiating model, b)the user's action model, c) the system's response model and d) the user's response model

was proposed and implemented into the first versions of a multimodal interactive system, evaluated by empirical studies with elderly persons, and then accordingly improved based on the evaluation data and implied suggestions. The final set of improved design guidelines were summarized (see the contributing publications in section 4.4) and have been used as the first fundamental aspect of the major focus of this research package ever since, especially for the development of the final version of the multimodal interactive system to be introduced in section 4.2.

4.1.2. The Unified Dialogue Model

As the second fundamental aspect of this research work, a unified dialogue model building on the unified dialogue modelling approach introduced in section 2.2 was developed, which a) is based on a generalized dialogue model (cf. [Ross et al., 2005]) for abstracting interaction by using illocutionary acts to represent discourse patterns in a recursive transition network, and b) integrates the classic information state update based management theory (cf. [Larsson and Traum, 2000]) to model discourse context as the attitudinal state of an intelligent agent for handling dynamic information exchange in context sensitive dialogues.
Figure 4.2 illustrates the unified dialogue model which comprises four transition network based sub-models regarding the four general transitions during the interaction process. Specifically,

- each interaction is initiated with *Dialogue(S, U)*, by the initialization of the system's

start state and a greeting request (i.e., S.request), which also triggers the initialization of the dialogue context with the information state update rule INIT.

- In *Dialogue(U, S)*, the dialogue proceeds with one of the user's instruction, request for certain information or restart action, which then triggers a transition to the system's further response or the restart of the dialogue. Accordingly, the information state is updated with the update rules INSTRUCT, REQUEST or INIT.

- After receiving user input, in *Response(S, U)*, the system tries to generate an appropriate response with respect to its current knowledge base and information state. There can be four returned responses: informing the user with requested data, rejecting an unsolvable request with or without certain reasons, providing choices for multiple options, or asking for further confirmation for taking critical actions, each of which triggers transitions to other sub-models and the corresponding information state update with update rules.

- Finally, *Response(U, S)* specifies that the user can conduct one of the four responses: accepting or rejecting the system's response, providing further or even new instructions or requests if the user wants to ignore the system's response, triggering further state transitions as well as information state updates.

With the generalized dialogue model based structure and the open definitions of the information state update rules, the unified dialogue model can be applied in various possible application scenarios in human-robot or human-computer interaction to support either single or multiple modalities featuring interaction. In this work package, this unified dialogue model is specified with the formal language CSP, implemented using the FormDia framework introduced in section 2.1, then integrated into a multimodal interactive system to support a flexible and context-sensitive, yet formally tractable and controllable multimodal interaction for elderly persons. More details about this model can be found in the contributing publications in section 4.4.

4.2. MIGSEP: the Multimodal Interactive Guidance System for Elderly Persons

According to the design and development foundation introduced in the previous sections, MIGSEP, the Multimodal Interactive Guidance System for Elderly Persons was developed. MIGSEP runs on a portable touch-screen tablet PC and is intended to be used as an interaction assistance system by an elderly or handicapped person seated in an electronic wheelchair that can navigate this person within complex environments autonomously.
The general architecture of MIGSEP is illustrated in figure 4.3. The *Unified Dialogue Manager* was developed based on the introduced unified dialogue model with the FormDia framework and functions as the central processing unit of the entire system to support a flexible and context-sensitive, yet formally controllable and extensible multimodal interaction management. On the one hand, an *Input Manager* interprets all incoming messages from the *GUI Action Recognizer* for GUI input events, the *Speech Recognizer* for natural language

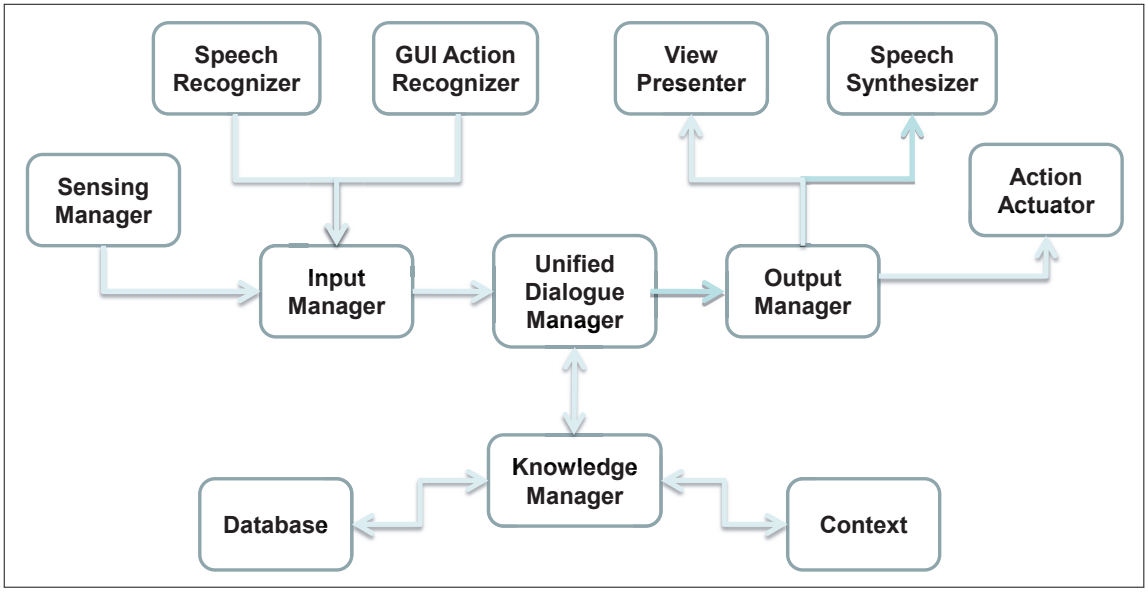


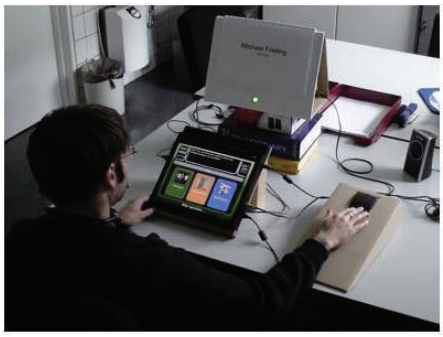
Figure 4.3.: The general architecture of MIGSEP

recognition and the *Sensing Manager*, for other possible sensor data messages from the electronic wheelchair, then sends them to the unified dialogue manager. On the other hand, an *Output Manager* handles all outgoing system messages and distributes them to the *View Presenter* for presenting visual feedback, the *Speech Synthesizer* to generate natural language responses and the *Action Actuator* to perform necessary motor actions, such as sending a driving request to the autonomous electronic wheelchair. The *Knowledge Manager*, constantly connected with the unified Dialogue Manager, can access static data from a *Database* component and process the dynamic information that is exchanged with the users during the interaction with a *Context* component.

All components of MIGSEP are connected via XML-based communication channels and each component can be treated as an open black box that can be accordingly implemented, modified or extended for specific domain, while the changes/extensions only cause changes in other components in the MIGSEP architecture on the data level. It provides a general open platform for both theoretical research and empirical studies on single- or multimodal interaction in different application domains or scenarios.

Based on the general architecture, the current MIGSEP system was implemented as a multimodal interactive guidance system to be used by elderly persons to navigate within the domain of hospital environments. Figure 4.4 shows the MIGSEP system interacting with a user. This MIGSEP system comprises a button device for triggering a press-to-talk signal, a green lamp to signalize the "being pressed and ready to talk" state, and the tablet PC, on which the MIGSEP system is running and the interface is displayed. The MIGSEP interface consists of two areas:

- Function-area contains the function button "start" on the top left for going to the start menu, the function button "toilet" below it concerns with the basic needs of elderly persons, and the text area besides them for displaying the system responses during the interaction;



Sys: <shows 3 cards, green for persons, yellow for rooms and blue for departments>
Hello, where do you want to go?
User: I want to go to a person named Wolf.
Sys: <shows a green bar texted with "Wolf", and 2 green cards, with photos, titles and names in the middle>
I found two persons named Wolf, to which person do you want to go?
User: he works in the cardiology department.
Sys: <enlarges the second green card, shows yes/no button>
Do you want to go to Mrs. Dr. Diana Wolf in the cardiology department?
User: <press the yes button>
Sys: <clears the screen and shows the 3 cards in the start menu again.>
OK, I have saved this goal. Where else do you want to go?

Figure 4.4.: Interaction with MIGSEP

- Choice-area displays information entities as single cards that can be selected, with a scrollbar indicating the position of the current displayed cards and a context sensitive coloured bar showing the current concerned context if necessary.

As an example, figure 4.4 shows the spoken language and touch-screen combined interaction between the MIGSEP system and a user who would like go to person named "Wolf" while there are two persons with the same first name "Wolf", and the MIGSEP system resolves the situation with context-sensitive multimodal interaction.

4.3. Empirical Evaluation of MIGSEP

In order to evaluate how well the MIGSEP system can assist elderly persons by focusing on the system usability as a whole, while regarding a comparison between the different input modalities: spoken language, gesture via touch-screen and the combination of both, as well as to enable continuously systematically and empirically improved development of multimodal interaction for elderly persons, a series of experimental studies were conducted with the elderly persons in the predefined age range and similar experiment settings. Concrete details about the experimental studies can be found in the contributing publications in section 4.4.
Since the empirical studies were focusing on multimodal interaction, how to evaluate multimodal data was another important issue to be dealt with in this research package. There has been a number of research studies investigated in developing metrics and frameworks for evaluating the performance of spoken dialogue systems (e.g., [Walker et al., 2000, Hajdinjak and Mihelic, 2006]); however, little research has been done with the evaluation of multimodal interactive systems, especially from the perspective of interaction. Therefore, based on Paradise (originally from [Walker et al., 1997]), a classic evaluation framework for spoken dialogue systems, an adapted version of a general evaluation framework for multimodal interaction was developed. Figure 4.5 illustrates the contribution of three essential factors to the overall usability of a multimodal interactive system: a) efficiency measure with the interaction related costs, b) subjective assessment with well-designed questionnaire, and c) task success with an AVT-based kappa coefficient. Specifically:

- The first factor: the measurement of efficiency is conducted based on the principles of the classic Paradise framework, by calculating all the possible objective data auto-

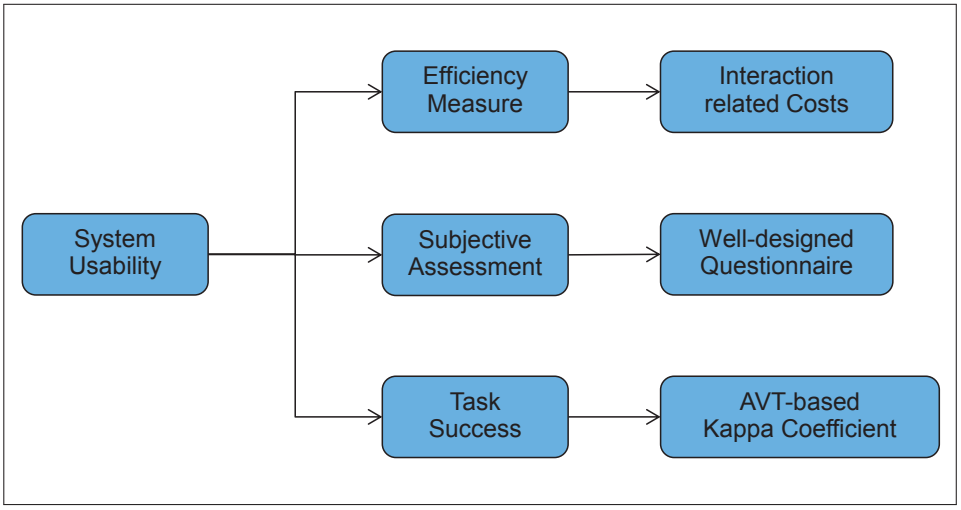


Figure 4.5.: An adapted framework to evaluate multimodal interaction

matically recorded during the multimodal interaction regarding essential interaction factors, such as user turns, system turns or elapsed time, etc. for each performed task.

- Regarding the second factor concerning with the subject user satisfaction of a multimodal interactive system, with the cooperation of the department of Medical Psychology and Medical Sociology at the University Medical Center Göttingen, evaluation questionnaires were especially designed with respect to the seven important aspects concerning multimodal interaction, i.e., system behaviour, speech output, textual output, interface presentation, task performing, user-friendliness and user perspective. These questionnaires can be used as general questionnaires to evaluate the subjective assessment of multimodal interactive systems.

- For the third factor, in the classic Paradise framework, a confusion matrix is needed for calculating the kappa coefficient to measure the effectiveness of task performance of a spoken dialogue system. However, an attribute value matrix (abbr. AVM), according to which a confusion matrix is constructed, cannot be built in the original way as the Paradise framework proposed, because it is not suitable for the data collected during multimodal interaction. In order to construct the needed confusion matrix, the concept of attribute value tree (abbr. AVT) was developed to replace AVM, where an AVT also contains all information to be exchanged during the multimodal interaction as AVM does for spoken language dialogue and therefore can deal with data recorded during multimodal interaction.
  In general, an AVT is defined as a finite state transition diagram, which uses all the expected correct ways, either touch-screen input or spoken language command, or other types of input events, as events to trigger state transitions within the transition diagram. E.g., in the simple AVT for the task "go to room 2602" illustrated in Figure 4.6 a), every correct interaction with the MIGSEP system can trigger transitions from the state [*MainView*], to e.g. [*RoomView*] by choosing the second card (*MS: select 1*

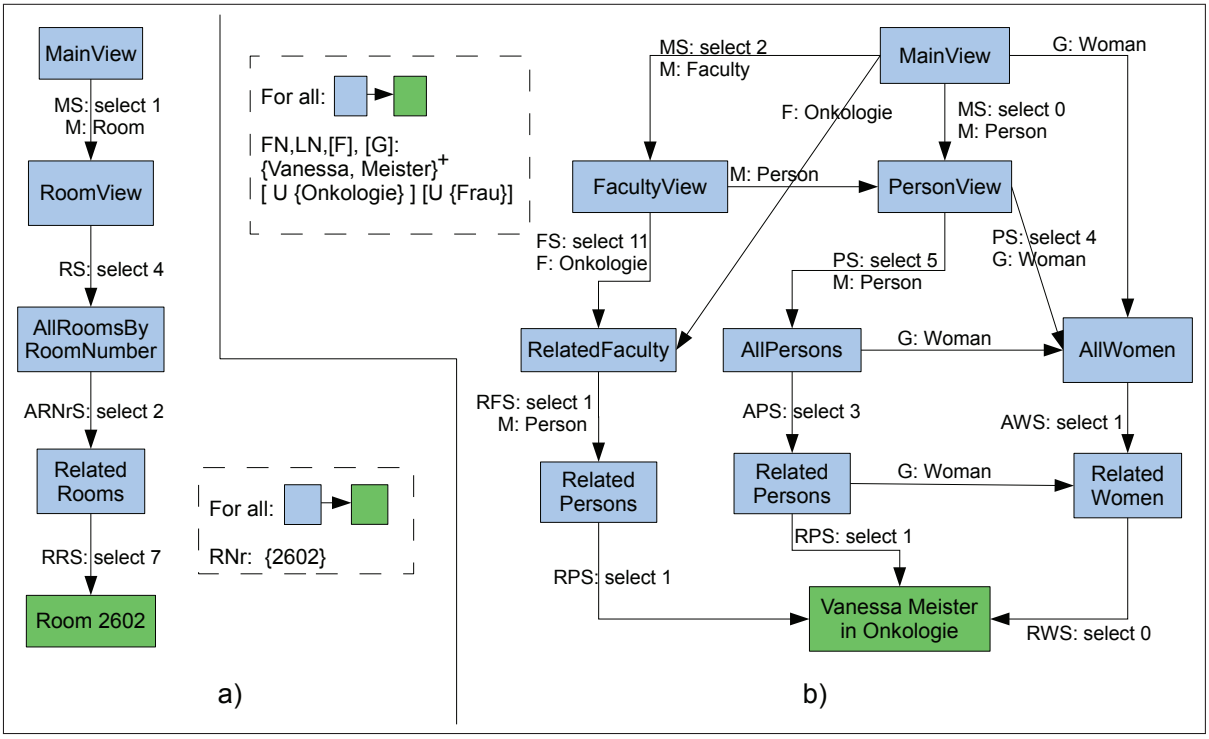


Figure 4.6.: Two Attribute Value Trees for the tasks a) "go to room 2602" and b) "go to Mrs. Vanessa Meister in the oncology department".

(counting from 0)), or using the spoken language command "I want to go to a room" (*M: Room*); then from [*RoomView*] to [*AllRoomsByRoomNumber*] with choosing the 5th card (*RS:select 4*); from [*AllRoomsByRoomNumber*] to [*RelatedRooms*] with choosing the 8th card (*RS: select 7*); or directly to the end state [*Room 2602*] from any state with the utterance "I want to go to room 2602" (*RNr: 2602*), etc. More complicated AVTs can also be constructed regarding the same principle, such as the AVT for the task "go to Mrs. Vanessa Meister in the oncology department" shown in figure 4.6 b).

As a result, a confusion matrix can be constructed by summarizing all multimodal data corresponding to one AVT. For example table 4.1 shows a confusion matrix for the simple task "go to room 2602", where "M" and "N" denote whether the actual data matched the expected attribute values in the AVT. E.g., there were 22 correctly performed actions [*MetaSelect (MS)*]; the [*AllRoomByRoomNumbersSelect (ARNrS)*] was wrongly performed 4 times and correctly 17 times; or the spoken language command regarding the [*room number (RNr)*] was mistakenly recognized by the system 16 times, etc. Note that, because of the width of the text, not all attributes of this confusion matrix can be shown in this example.

Given a confusion matrix, the Kappa coefficient can be calculated with:

$$k = \frac{P(A)-P(E)}{1-P(E)} \text{ with } P(A) = \frac{\sum_{i=1}^{n} M(i,M)}{T},\ P(E) = \sum_{i=1}^{n} (\frac{M(i)}{T})^2 \text{ ([Walker et al., 1997])}$$

where *P(A)* is the proportion of times that the actual data agreed with the expected attribute values, *P(E)* is the proportion of times that the actual data are expected to be agreed on by chance, *M(i, M)* is the value of the matched cell of row *i*, *M(i)* the

Table 4.1.: A sample confusion matrix for the task "go to room 2602"

Data	M	N	M	N	...	M	N	Sum
MS	22							22
ARNrS			17	4				21
...								...
RNr						93	16	109

sum of the cells of row i, and T the sum of all cells.
Finally, the higher a value of the Kappa coefficient is, the higher is the effectiveness of task performance for the concerned multimodal interaction.

The empirical data of the series of empirical studies for evaluating MIGSEP used by elderly persons were all evaluated with the adapted framework. The overall positive results showed high effectiveness of task performance, high efficiency of interaction and high user satisfaction with the implemented MIGSEP system. The concrete details about the analysis and discussion of the results can be found in the contributing publications in section 4.4.

4.4. Contribution of the Corresponding Publications

The major research work of this package has been concentrating on multimodal interaction in the AAL context for elderly persons by taking ageing-centered characteristics into account. The contributed work comprises two important aspects: a) the design, development and implementation of multimodal interaction for elderly persons, and b) the model-based empirical evaluation of an elderly-friendly multimodal interactive guidance system. Specifically:

- according to the traditional design principles of conventional interactive systems and the study of most common declining processes in sensory, perceptual, motor and cognitive abilities during normal ageing, [10] proposed a list of design guidelines for developing multimodal interactive systems for elderly persons. A pilot study was then conducted to test a prototype of an interactive system with touch-screen interface that implemented the design guidelines. The positive results provided support for the proposed guidelines, with implied suggestions.

- with the formal language based framework and the unified dialogue modelling approach, a unified dialogue model was developed and implemented into an interactive guidance system to support a flexible and context-sensitive, yet formally tractable and controllable interaction for elderly persons ([7]). The spoken natural language interface of the system was then tested by an experiment with 16 elderly persons, which again provided positive results and advised further improvements.

- with the revised design guidelines for the multimodal interaction for elderly persons and the unified dialogue modelling and management, the touch-screen interface and the spoken language interface of an improved interactive guidance system were tested and compared with 31 elderly persons in an empirical study ([4]). The overall positive

results of both modalities confirmed the proposed guidelines, approaches and frameworks, yet found no significant evidence for one preferred modality, which implies the necessity of studying the combination of both.

- based on the data of the previous studies, the proposed elderly-centered guidelines, approaches and frameworks, the multimodal interactive guidance system for elderly persons was developed with both touch-screen input and spoken language input, then tested with an elaborated experimental study with 33 elderly persons ([5]). An adapted version of a general evaluation framework was also developed and proposed to evaluate the data of the multimodal interaction.

- As the summary of the empirical studies, [2] reported a detailed comparison of the touch-screen, spoken language inputs and the combination of both, and highlighted the assessment of the complete multimodal interactive guidance system concerning its effectiveness of task performance, efficiency of interaction and user satisfaction. The positive result of the analysis again validated the research effort based on the proposed elderly-centered design guidelines, the especially developed unified dialogue model and the supporting frameworks.

Furthermore, the author has also contributed to some other research work in the field of multimodal interaction in AAL environments, e.g., [15] reported an empirical study on identifying the behavioural patterns of users with physical disabilities while driving an electronic wheelchair with a safety assistant, which advised further improvement of the driving assistance system, or [8] presented the multimodal interaction in the Bremen Ambient Assisted Living Lab (BAALL, [Krieg-Brückner, 2013]), which combines the speech interface and the gesture interface to control multiple devices within BAALL, or in [6], empirical studies were conducted and reported, on using speech and gesture to interact with the devices in BAALL, which provided interesting results and motivated the development of multimodal interactive systems in similar ambient assisted living environments.

4.5. Possible Future Work

The presented work continued the pursuit of the goal of building effective, efficient, user-friendly, adaptive and robust multimodal interactive systems and framework in ambient assisted living environments. Therefore, further research work can be carried out with respect to many related aspects, such as:

- Supervised and reinforcement learning techniques can be applied to improve the information state update mechanism within the unified dialogue model for supporting an adaptive user interface that features user-tailored yet still elderly-friendly interaction.

- With corresponding adaptation on only the data level, the current MIGSEP architecture and system based on the especially developed design guidelines and the general unified dialogue model can also be extended to support further application scenarios, such as multimodal interaction within a smart home environment for controlling the fully equipped technological infrastructure (e.g., BAALL ([Krieg-Brückner, 2013])), as well as the evaluation of similar multimodal interaction.

- Although the complementary input modalities are not suitable for elderly persons due to the reported decline of human abilities, they are more applicable for persons within other age ranges in many other application domains (e.g., [Lalanne et al., 2009]). The current unified dialogue model is built at the illocutionary level, therefore, a fusion engine can be developed for interpreting the combination of multimodal input events and delivering the illocutionary leveled semantic representation that can then be directly used in the current interaction model, without causing changes on the other components of the MIGSEP architecture.

Full List of Publications by the Author

[1] Cui Jian and Hui Shi. A conceptual model for human-robot collaborative spatial navigation. In *The Tenth Asia-Pacific Conference on Conceptual Modelling (APCCM2014)*, 2014. accepted.

[2] Cui Jian, Hui Shi, Frank Schafmeister, Carsten Rachuy, Nadine Sasse, Holger Schmidt, Volker Hoemberg, and Nicole von Steinbüchel. Modality preference in multimodal interaction for elderly persons. In *Biomedical Engineering Systems and Technologies 2013*, Lecture Notes, Communications in Computer and Information Science (CCIS). Springer-Verlag, 2014. to appear.

[3] Cui Jian and Hui Shi. Resolving conceptual mode confusion with qualitative spatial knowledge in human-robot interaction. In *Proceedings of the 11th International Conference on Spatial Information Theory*, Lecture Notes in Computer Science (LNCS), pages 91–108. Springer-Verlag, 2013.

[4] Cui Jian, Hui Shi, Frank Schafmeister, Carsten Rachuy, Nadine Sasse, Holger Schmidt, Volker Hoemberg, and Nicole von Steinbüchel. Touch and speech: Multimodal interaction for elderly persons. In *Biomedical Engineering Systems and Technologies 2012*, Lecture Notes, Communications in Computer and Information Science (CCIS), pages 385–400. Springer-Verlag, 2013.

[5] Cui Jian, Hui Shi, Frank Schafmeister, Carsten Rachuy, Nadine Sasse, Holger Schmidt, and Nicole von Steinbüchel. Better choice? combing touch and speech in multimodal interaction for elderly persons. In *6th International Conference on Health Informatics*, 2013. to appear.

[6] Dimitra Anastasiou, Cui Jian, and Christoph Stahl. A german-chinese speech-gesture corpus of device control in a smart home. In *Proceedings of The 6th Workshop on Affect and Behaviour Related Assistance, the 6th International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 2013. to appear.

[7] Cui Jian, Frank Schafmeister, Carsten Rachuy, Nadine Sasse, Hui Shi, Holger Schmidt, and Nicole von Steinbüchel. Evluating a spoken language interface of a multimodal interactive guidance system for elderly persons. In *5th International Conference on Health Informatics*, pages 87–96. SciTePress, 2012.

[8] Dimitra Anastasiou, Cui Jian, and Desislava Zhekova. Speech and gesture interaction in an ambient assisted living lab. In *Proceedings of the 1st Workshop on Speech and Multimodal Interaction in Assistive Environments*, SMIAE '12, pages 18–27, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics (ACL).

[9] Vivien Mast, Cui Jian, and Desislava Zhekova. Elaborate descriptive information in indoor route instructions. In *The 34th annual meeting of the Cognitive Science Society*, pages 1972–1977, 2012.

[10] Cui Jian, Frank Schafmeister, Carsten Rachuy, Nadine Sasse, Hui Shi, Holger Schmidt, and Nicole von Steinbüchel. Towards effective, efficient and elderly-friendly multimodal interaction. In *The 4th International Conference on PErvasive Technologies Related to Assistive Environments*, pages 45:1–45:8, New York, USA, 2011. ACM.

[11] Hui Shi, Cui Jian, and Carsten Rachuy. Evaluation of a unified dialogue model for human-computer interaction. *International Journal of Computational Linguistics and Applications*, 2(1):155–173, 2011.

[12] Carsten P.J. Gondorf and Cui Jian. Supporting inferences in space - a wayfinding task in a multilevel building. In *The 2nd Workshop on Computational Models of Spatial Language Interpretation and Generation*, pages 48–52, 2011.

[13] Cui Jian, Desislava Zhekova, Hui Shi, and John Bateman. Deep reasoning in clarification dialogues with mobile robots. In *19th European Conference on Artificial Intelligence (ECAI 2010)*, pages 177–182, Amsterdam, 2010. IOS Press.

[14] Hui Shi, Cui Jian, and Bernd Krieg-Brückner. Qualitative spatial modelling of human route instructions to mobile robots. In *The Third International Conference on Advances in Computer-Human Interactions*, pages 1–6. IEEE, 2010.

[15] Carsten Fischer, Hui Shi, Cui Jian, Frank Schafmeister, Nils Menrad, Nicole V. Steinbüchel, Kerstin Schill, and Bernd Krieg-Brükner. Modelling user behaviour while driving an intelligent wheelchair. In *3rd International Conference on Health Informatics*, pages 330–336, 2010.

[16] Cui Jian, Hui Shi, and Bernd Krieg-Brückner. Simspace: A tool to interpret route instructions with qualitative spatial knowledge. In *AAAI Spring Symposium on Benchmarking of Qualitative Spatial and Temporal Reasoning Systems*, pages 47–48, 2009.

[17] Bernd Krieg-Brückner, Hui Shi, Carsten Fischer, Thomas Röfer, Cui Jian, and Kerstin Schill. What safety assistance is needed for wheelchair drivers? In *2. Deutscher AAL-Kongress*, Berlin, 2009. VDE-Verlag.

Bibliography

[Alston, 2000] Alston, W. P. (2000). *Illocutionary Acts and Sentence Meaning*. PhD thesis, Cornell University, Cornell University Press.

[Becker et al., 2009] Becker, E., Le, Z., Park, K., Lin, Y., and Makedon, F. (2009). Event-based experiments in an assistive environment using wireless sensor networks and voice recognition. In *Proceedings of the 2nd International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '09, pages 17:1–17:8, New York, NY, USA. ACM.

[Bellotto et al., 2013] Bellotto, N., Hanheide, M., and Van de Weghe, N. (2013). Qualitative design and implementation of human-robot spatial interactions. In *Proceedings of International Conference on Social Robotics (ICSR)*, page In Press.

[Bhatt et al., 2011] Bhatt, M., Lee, J. H., and Schultz, C. (2011). Clp(qs): a declarative spatial reasoning framework. In *Proceedings of the 10th international conference on Spatial information theory*, COSIT'11, pages 210–230, Berlin, Heidelberg. Springer-Verlag.

[BKB, 2005] BKB, A. (2005). uDraw(graph) - the powerful solution for graph visualization.

[Boll et al., 2010] Boll, S., Heuten, W., Meyer, E. M., and Meis, M. (2010). Development of a multimodal reminder system for older persons in their residential home. *Informatics for Health and Social Care*, 35(3-4):104–124.

[Bugmann et al., 2004] Bugmann, G., Klein, E., Lauria, S., and Kyriacou, T. (2004). Corpus-based robotics: A route instruction example. In *Proceedings of the 8th International Conference on Intelligent Autonomous Systems*, pages 96–103.

[Burbeck, 1987] Burbeck, S. (1987). Applications programming in smalltalk-80(tm): How to use model-view-controller (mvc).

[Burke and Mackay, 1997] Burke, D. M. and Mackay, D. G. (1997). Memory, language and ageing. *Philosophical Transactions of the Royal Society B: Bilogoical Sciences*, 352(1363):1845–1856.

[Charness and Boot, 2009] Charness, N. and Boot, W. R. (2009). Aging and Information Technology Use: Potential and Barriers. *Current Directions in Psychological Science*, 18(5):253–258.

[Chu-Carroll, 1999] Chu-Carroll, J. (1999). Form-based reasoning for mixed-initiative dialogue management in information-query systems. In *EUROSPEECH*. ISCA.

[Cohn et al., 1997] Cohn, A. G., Bennett, B., Gooday, J., and Gotts, N. M. (1997). Qualitative spatial representation and reasoning with the region connection calculus. *Geoinformatica*, 1(3):275–316.

[Craik and Jennings, 1992] Craik, F. I. M. and Jennings, J. M. (1992). Human memory. In Craik, F. I. M. and Salthouse, T. A., editors, *The handbook of aging and cognition*, pages 51–110. Erlbaum, Hillsdale, NJ.

[D'Andrea et al., 2009] D'Andrea, A., D'Ulizia, A., Ferri, F., and Grifoni, P. (2009). A multimodal pervasive framework for ambient assisted living. In *Proceedings of the 2nd International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '09, pages 39:1–39:8, New York, NY, USA. ACM.

[Denis, 1997] Denis, M. (1997). The Description of Routes: A Cognitive Approach to the Production of Spatial Discourse. *Cahiers Psychologie Cognitive*, 16(4):409–458.

[D'Ulizia et al., 2007] D'Ulizia, A., Ferri, F., and Grifoni, P. (2007). A hybrid grammar-based approach to multimodal languages specification. In *Proceedings of the 2007 OTM confederated international conference on On the move to meaningful internet systems - Volume Part I*, OTM'07, pages 367–376, Berlin, Heidelberg. Springer-Verlag.

[Dylla et al., 2006] Dylla, F., Frommberger, L., Wallgrün, J. O., and Wolter, D. (2006). SparQ: A toolbox for qualitative spatial representation and reasoning. In *Proceedings of the Workshop on Qualitative Constraint Calculi: Application and Integration (KI 2006)*, pages 79–90.

[(Europe), 2010] (Europe), F. S. (2010). Home of the fdr2 model-checker and other csp tools.

[Fisk et al., 2009] Fisk, A. D., Rogers, W. A., Charness, N., Czaja, S. J., and Sharit, J. (2009). *Designing for older adults: Principles and creative human factors approaches*. Boca Raton: CRC Press, 2 edition.

[Fozard and Gordon-Salant, 2001] Fozard, J. L. and Gordon-Salant, S. (2001). Changes in vision and hearing with aging. *Handbook of the psychology of aging*, pages 241–266.

[Frank, 1996] Frank, A. U. (1996). Qualitative spatial reasoning: Cardinal directions as an example. *International Journal of Geographical Information Science*, 10(3):269–290.

[Freksa, 1992] Freksa, C. (1992). Using orientation information for qualitative spatial reasoning. In *Proceedings of the International Conference GIS - From Space to Territory: Theories and Methods of Spatio-Temporal Reasoning on Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, pages 162–178, London, UK, UK. Springer-Verlag.

[Galindo et al., 2005] Galindo, C., Saffiotti, A., Coradeschi, S., Buschka, P., Fernandez-Madrigal, J., and Gonzalez, J. (2005). Multi-hierarchical semantic maps for mobile robotics. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-05)*, pages 2278–2283, Edmonton, Canada.

[Goetze et al., 2012] Goetze, S., Fischer, S., Moritz, N., Appell, J.-E., and Wallhoff, F. (2012). Multimodal human-machine interaction for service robots in home-care environments. In *Proceedings of the 1st Workshop on Speech and Multimodal Interaction in Assistive Environments*, SMIAE '12, pages 1–7, Stroudsburg, PA, USA. Association for Computational Linguistics.

[Goetze et al., 2010] Goetze, S., Moritz, N., Appell, J., Meis, M., Bartsch, C., and Bitzer, J. (2010). Acoustic user interfaces for ambient assisted living technologies. *Informatics for Health and Social Care, SI Ageing and Technology*, vol. 35, no. 4.

[Hajdinjak and Mihelic, 2006] Hajdinjak, M. and Mihelic, F. (2006). The paradise evaluation framework: Issues and findings. *Comput. Linguist.*, 32(2):263–272.

[Hawthorn, 2007] Hawthorn, D. (2007). Interface design and engagement with older people. *Behaviour and IT*, 26(4):333–341.

[Heise.de, 2009] Heise.de (2009). Uni bremen: Sprechender kaffeeautomat gibt auch auskünfte.

[Hirtle, 2008] Hirtle, S. C. (2008). Landmarks for navigation in human and robots. In Jefferies, M. E. and Yeap, W.-K., editors, *Robotics and Cognitive Approaches to Spatial Mapping*, volume 38 of *Springer Tracts in Advanced Robotics*, chapter 8, pages 203–214. Springer Berlin Heidelberg.

[Hoare, 1978] Hoare, C. A. R. (1978). Communicating sequential processes. *Commun. ACM*, 21(8):666–677.

[Ivanecky et al., 2011] Ivanecky, J., Mehlhase, S., and Mieskes, M. (2011). An intelligent house control using speech recognition with integrated localization. In *Ambient Assisted Living, 4. AAl-congree 2011, Berlin, Germany*, pages 51–62. Springer Berlin Heidelberg.

[Kline and Scialfa, 1997] Kline, D. and Scialfa, C. (1997). Sensory and perceptual functioning: Basic research and human factors implications. In *Handbook of Human Factors and the Older Adult*, pages 27–54. Academic Press, New York.

[Kollar et al., 2010] Kollar, T., Tellex, S., Roy, D., and Roy, N. (2010). Toward understanding natural language directions. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, HRI '10, pages 259–266, Piscataway, NJ, USA. IEEE Press.

[Kotary and Hoyer, 1995] Kotary, L. and Hoyer, W. J. (1995). Age and the Ability To Inhibit Distractor Information in Visual Selective Attention. *Experimental Aging Research: An International Journal Devoted to the Scientific Study of the Aging Process*, 21(2):159–171.

[Koulouri and Lauria, 2009] Koulouri, T. and Lauria, S. (2009). A corpus-based analysis of route instructions in human-robot interaction. In *Towards Autonomous Robotic Systems (TAROS)*, pages 281–288. Londonderry, University of Ulster.

[Krieg-Brückner, 2013] Krieg-Brückner, B. (2013). BAALL - Bremen Ambient Assisted Living Lab.

[Krieg-Brückner et al., 2004] Krieg-Brückner, B., Frese, U., Lüttich, K., Mandel, C., Mossakowski, T., and Ross, R. J. (2004). Specification of an ontology for route graphs. In Freksa, C., Knauff, M., Krieg-Brückner, B., Nebel, B., and Barkowsky, T., editors, *Spatial Cognition*, volume 3343 of *Lecture Notes in Computer Science*, pages 390–412. Springer.

[Krieg-Brückner and Shi, 2006] Krieg-Brückner, B. and Shi, H. (2006). Orientation calculi and route graphs: Towards semantic representations for route descriptions. In Raubal, M., Miller, H., Frank, A., and Goodchild, M., editors, *Geographic Information Science - Fourth International Conference, GIScience 2006*, volume 4197 of *Lecture Notes in Computer Science*. Springer; http://www.springer.de.

[Kurata, 2008] Kurata, Y. (2008). The 9 + -intersection: A universal framework for modeling topological relations. In *Proceedings of the 5th international conference on Geographic Information Science*, GIScience '08, pages 181–198, Berlin, Heidelberg. Springer-Verlag.

[Lalanne et al., 2009] Lalanne, D., Nigay, L., Palanque, p., Robinson, P., Vanderdonckt, J., and Ladry, J.-F. (2009). Fusion engines for multimodal input: a survey. In *Proceedings of the 2009 international conference on Multimodal interfaces*, ICMI-MLMI '09, pages 153–160, New York, NY, USA. ACM.

[Lamel et al., 1999] Lamel, L., Rosset, S., Gauvain, J. L., and Bennacef, S. (1999). The limsi arise system for train travel information. In *Proceedings of the Acoustics, Speech, and Signal Processing, 1999. on 1999 IEEE International Conference*, ICASSP '99, pages 501–504, Washington, DC, USA. IEEE Computer Society.

[Lankenau and Röfer, 2000] Lankenau, A. and Röfer, T. (2000). Smart Wheelchairs - State of the Art in an Emerging Market. *Künstliche Intelligenz. Schwerpunkt Autonome Mobile Systeme*, 4:37–39.

[Lankenau and Röfer, 2001] Lankenau, A. and Röfer, T. (2001). A Safe and Versatile Mobility Assistant. *Reinventing the Wheelchair. IEEE Robotics and Automation Magazine*, pages 27–37.

[Larsson and Traum, 2000] Larsson, S. and Traum, D. R. (2000). Information state and dialogue management in the trindi dialogue move engine toolkit. *Nat. Lang. Eng.*, 6(3-4):323–340.

[Lauria et al., 2002] Lauria, S., Kyriacou, T., Bugmann, G., Bos, J., and Klein, E. (2002). Converting natural language route instructions into robot executable procedures. In *Proceedings of the 2002 IEEE International Workshop on Human and Robot Interactive Communication*, pages 223–228.

[Lecoeuche, 2001] Lecoeuche, R. (2001). Learning optimal dialogue management rules by using reinforcement learning and inductive logic programming. In *NAACL '01: Second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies 2001*, pages 1–7, Morristown, NJ, USA. Association for Computational Linguistics.

[Lemon and Liu, 2006] Lemon, O. and Liu, X. (2006). Dude: a dialogue and understanding development environment, mapping business process models to information state update dialogue systems. In *Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics: Posters & Demonstrations*, EACL '06, pages 99–102, Stroudsburg, PA, USA. Association for Computational Linguistics.

[Leung et al., 2012] Leung, R., Tang, C., Haddad, S., Mcgrenere, J., Graf, P., and Ingriany, V. (2012). How older adults learn to use mobile devices: Survey and field investigations. *ACM Trans. Access. Comput.*, 4(3):11:1–11:33.

[Li et al., 2009] Li, L., Williams, J. D., and Balakrishnan, S. (2009). Reinforcement Learning for Dialog Management using Least-Squares Policy Iteration and Fast Feature Selection. In *Interspeech*, Brighton.

[Mackay and James, 2004] Mackay, D. G. and James, L. E. (2004). Sequencing, speech production, and selective effects of aging on phonological and morphological speech errors. *Psychol Aging*, 19(1):93–107.

[Mandel et al., 2009] Mandel, C., Lüth, T., Laue, T., Röfer, T., Gräser, A., and Krieg-Brückner, B. (2009). Navigating a smart wheelchair with a brain-computer interface interpreting steady-state visual evoked potentials. In *Proceedings of the 2009 IEEE/RSJ international conference on Intelligent robots and systems*, IROS'09, pages 1118–1125, Piscataway, NJ, USA. IEEE Press.

[Marge and Rudnicky, 2010] Marge, M. and Rudnicky, A. I. (2010). Comparing spoken language route instructions for robots across environment representations. In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, SIGDIAL '10, pages 157–164, Stroudsburg, PA, USA. Association for Computational Linguistics.

[Martínez Mozos, 2010] Martínez Mozos, O. (2010). Conceptual spatial representation of indoor environments. In *Semantic Labeling of Places with Mobile Robots, Springer Tracts in Advanced Robotics*, pages 83–97. Springer Berlin Heidelberg.

[McDowd and Craik, 1988] McDowd, J. M. and Craik, F. I. (1988). Effects of aging and task difficulty on divided attention performance. *Journal of experimental psychology. Human perception and performance*, 14(2):267–280.

[McTear, 1998] McTear, M. (1998). Modelling spoken dialogues with state transition diagrams: Experiences with the CSLU toolkit. In *ICSLP-98*, volume 4, pages 1223–1226, Sydney.

[Möller et al., 2008] Möller, S., Gödde, F., and Wolters, M. (2008). Corpus analysis of spoken smart-home interactions with older users. In *LREC*. European Language Resources Association.

[Nazemi et al., 2011] Nazemi, K., Burkhardt, D., Stab, C., Breyer, M., Wichert, R., and Fellner, D. W. (2011). Natural gesture interaction with accelerometer-based devices in

ambient assisted environments. In Wichert, R. and Eberhardt, B., editors, *Ambient Assisted Living, 4. AAL-Kongress 2011*, Advanced Technologies and Societal Change, pages 75–90, Berlin, Heidelberg, Germany. VDE, Springer.

[Nesselrath et al., 2011] Nesselrath, R., Lu, C., Schulz, C. H., Frey, J., and Alexandersson, J. (2011). A gesture based system for context-sensitive interaction with smart homes. In Wichert, R. and Eberhardt, B., editors, *Ambient Assisted Living, 4. AAL-Kongress 2011*, Advanced Technologies and Societal Change, pages 209–219, Berlin, Heidelberg, Germany. VDE, Springer.

[Pappu and Rudnicky, 2012] Pappu, A. and Rudnicky, A. (2012). The structure and generality of spoken route instructions. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, SIGDIAL '12, pages 99–107, Stroudsburg, PA, USA. Association for Computational Linguistics.

[Peckham, 1993] Peckham, J. (1993). A new generation of spoken dialogue systems: results and lessons from the sundial project. In *EUROSPEECH*. ISCA.

[Pietquin et al., 2011] Pietquin, O., Geist, M., Chandramohan, S., and Frezza-Buet, H. (2011). Sample-efficient batch reinforcement learning for dialogue management optimization. *ACM Trans. Speech Lang. Process.*, 7(3):7:1–7:21.

[Reason, 1990] Reason, J. (1990). *Human Error*. Cambridge University Press.

[Röfer and Lankenau, 2002] Röfer, T. and Lankenau, A. (2002). Route-Based Robot Navigation. *Künstliche Intelligenz - Themenheft Spatial Cognition*, pages 29–31.

[Röfer et al., 2009] Röfer, T., Laue, T., and Gersdorf, B. (2009). iWalker - An Intelligent Walker providing Services for the Elderly. *Technically Assisted Rehabilitation 2009*.

[Roger et al., 2007] Roger, M., Bonnardel, N., and Le Bigot, L. (2007). Spatial cognition in a navigation task: effects of initial knowledge of an environment and spatial abilities on route description. In *Proceedings of the 14th European conference on Cognitive ergonomics: invent! explore!*, ECCE '07, pages 237–242, New York, NY, USA. ACM.

[Roscoe, 1994] Roscoe, A. W. (1994). Model-checking csp. In Roscoe, A. W., editor, *A Classical Mind: essays in Honour of C.A.R. Hoare*, pages 353–378. Prentice Hall International (UK) Ltd., Hertfordshire, UK, UK.

[Roscoe et al., 1997] Roscoe, A. W., Hoare, C. A. R., and Bird, R. (1997). *The Theory and Practice of Concurrency*. Prentice Hall PTR, Upper Saddle River, NJ, USA.

[Ross et al., 2005] Ross, R., Bateman, J., and Shi, H. (2005). Using generalised dialogue models to constrain information state based dialogue systems. In *Proceedings of the Symposium on Dialogue Modelling and Generation*, pages 1–8.

[Schieber, 1992] Schieber, F. (1992). Aging and the senses. In Cohen, G. D., Sloane, R. B., Lebowitz, B. D., Wykle, M., Hooyman, N. R., and Birren, J. E., editors, *Handbook of Mental Health and Aging*, volume 2, pages 251–306. Academic Press, New York.

[Schultz et al., 2006] Schultz, C. P. L., Guesgen, H. W., and Amor, R. (2006). Computer-human interaction issues when integrating qualitative spatial reasoning into geographic information systems. In *Proceedings of the 7th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: design centered HCI*, CHINZ '06, pages 43–51, New York, NY, USA. ACM.

[Seneff and Polifroni, 2000] Seneff, S. and Polifroni, J. (2000). Dialogue management in the mercury flight reservation system. In *Proceedings of the 2000 ANLP/NAACL Workshop on Conversational systems*, ANLP/NAACL-ConvSyst '00, pages 11–16, Stroudsburg, PA, USA. Association for Computational Linguistics.

[Shi and Bateman, 2005] Shi, H. and Bateman, J. (2005). Developing human-robot dialogue management formally. In *Proceedings of Symposium on Dialogue Modelling and Generation*, Amsterdam, Netherlands.

[Shi and Krieg-Brückner, 2008] Shi, H. and Krieg-Brückner, B. (2008). Modelling route instructions for robust human-robot interaction on navigation tasks. *International Journal of Software and Informatics*, 2(1):33–60. ISSN: 1673-7288.

[Shi et al., 2006] Shi, H., Mandel, C., and Ross, R. J. (2006). Interpreting route instructions as qualitative spatial actions. In Barkowsky, T., Knauff, M., Ligozat, G., and Montello, D. R., editors, *Spatial Cognition*, volume 4387 of *Lecture Notes in Computer Science*, pages 327–345. Springer.

[Shi et al., 2005] Shi, H., Ross, R. J., and Bateman, J. (2005). Formalising control in robust spoken dialogue systems. In *Proceedings of the Third IEEE International Conference on Software Engineering and Formal Methods*, SEFM '05, pages 332–341, Washington, DC, USA. IEEE Computer Society.

[Shi et al., 2010] Shi, H., Ross, R. J., Tenbrink, T., and Bateman, J. (2010). Modelling illocutionary structure: combining empirical studies with formal model analysis. In *Proceedings of the 11th international conference on Computational Linguistics and Intelligent Text Processing*, CICLing'10, pages 340–353, Berlin, Heidelberg. Springer-Verlag.

[Shi and Tenbrink, 2009] Shi, H. and Tenbrink, T. (2009). Telling rolland where to go: Hri dialogues on route navigation. In Conventry, K. R., Tenbrink, T., and Bateman, J., editors, *Spatial Language and Dialogue (Explorations in Language and Space)*, pages 177–216. Oxford University Press.

[Sitter and Stein, 1992] Sitter, S. and Stein, A. (1992). Modeling the illocutionary aspects of information-seeking dialogues. *Inf. Process. Manage.*, 28(2):165–180.

[Smith et al., 1999] Smith, M. W., Sharit, J., and Czaja, S. J. (1999). Aging, motor control, and the performance of computer mouse tasks. *Human Factors*, 41(3):389–396.

[Stoltzfus et al., 1996] Stoltzfus, E. R., Hasher, L., and Zacks, R. T. (1996). Working memory and aging: Current status of the inhibitory view. *Counterpoints in Cognition: Working Memory and Human Cognition*, pages 66–68.

[Takahashi et al., 2003] Takahashi, S., Morimoto, T., Maeda, S., and Tsuruta, N. (2003). Dialogue experiment for elderly people in home health care system. In Matousek, V. and Mautner, P., editors, *TSD*, volume 2807 of *Lecture Notes in Computer Science*, pages 418–423. Springer.

[Traum and Larsson, 2003] Traum, D. and Larsson, S. (2003). The information state approach to dialogue management. In *Current and New Directions in Discourse and Dialogue*, pages 325–353. Springer.

[Tversky and Lee, 1998] Tversky, B. and Lee, P. U. (1998). How space structures language. In *Spatial Cognition, An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*, pages 157–176, London, UK, UK. Springer-Verlag.

[Varges et al., 2008] Varges, S., Riccardi, G., and Quarteroni, S. (2008). Persistent information state in a data-centric architecture. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, SIGdial '08, pages 68–71, Stroudsburg, PA, USA. Association for Computational Linguistics.

[Vasudevan et al., 2007] Vasudevan, S., Gächter, S., Nguyen, V., and Siegwart, R. (2007). Cognitive maps for mobile robots - an object based approach. *Robot. Auton. Syst.*, 55(5):359–371.

[Walker et al., 2000] Walker, M., Kamm, C., and Litman, D. (2000). Towards developing general models of usability with paradise. *Nat. Lang. Eng.*, 6(3-4):363–377.

[Walker et al., 1997] Walker, M. A., Litman, D. J., Kamm, C. A., and Abella, A. (1997). Paradise: a framework for evaluating spoken dialogue agents. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, ACL '98, pages 271–280, Stroudsburg, PA, USA. Association for Computational Linguistics.

[Wallgrün et al., 2007] Wallgrün, J. O., Frommberger, L., Wolter, D., Dylla, F., and Freksa, C. (2007). Qualitative spatial representation and reasoning in the sparq-toolbox. In *Proceedings of the 2006 international conference on Spatial Cognition V: reasoning, action, interaction*, SC'06, pages 39–58, Berlin, Heidelberg. Springer-Verlag.

[Wang et al., 2013] Wang, T., Shi, H., and Krieg-Brückner, B. (2013). Using observations to derive cardinal direction relations between regions. In *Proceedings of the 27th International Workshop on Qualitative Reasoning*, pages 139–145.

[Werner et al., 2000] Werner, S., Krieg-Brückner, B., and Herrmann, T. (2000). Modelling navigational knowledge by route graphs. In *Spatial Cognition II, Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications*, pages 295–316, London, UK, UK. Springer-Verlag.

[Werner et al., 1997] Werner, S., Krieg-Brückner, B., Mallot, H. A., Schweizer, K., and Freksa, C. (1997). Spatial cognition: The role of landmark, route, and survey knowledge in human and robot navigation. In *Ph.D. in Computer Science in*, pages 41–50. Springer.

[Zender et al., 2008] Zender, H., Martínez Mozos, O., Jensfelt, P., Kruijff, G. J. M., and Burgard, W. (2008). Conceptual spatial representations for indoor mobile robots. *Robot. Auton. Syst.*, 56(6):493–502.

[Ziefle and Bay, 2005] Ziefle, M. and Bay, S. (2005). How older adults meet complexity: Aging effects on the usability of different mobile phones. *Behaviour and IT*, 24(5):375–389.

[Zue et al., 2000] Zue, V., Seneff, S., Glass, J., Polifroni, J., Pao, C., Hazen, T. J., and Hetherington, L. (2000). Jupiter: A telephone-based conversational interface for weather information. *IEEE Trans. on Speech and Audio Processing*, 8:85–96.

Appendix A.

Accumulated Publications

The author has contributed to 17 publications in international workshops, conferences and journals (see the list of publications by the author in the previous section) and 11 of them are selected to contribute to the cumulative doctoral thesis.
Before listing all the accumulated publications in this chapter, the contribution of the author to each of the accumulated publications is given as a tuple that consists of 3 percentage numbers (a, b, c), indicating the author's individual contribution to a) the concept of the work, b) the content of the paper or the implemented work, and c) the writing of the paper, as follows:

1. (80%, 80%, 70%) **Jian**, Shi and Krieg-Brückner: SimSpace: A Tool to Interpret Route Instructions with Qualitative Spatial Knowledge. In: AAAI Symposium on Benchmarking of Qualitative Spatial and Temporal Reasoning Systems, 2009.

2. (20%, 80%, 20%) Shi, **Jian** and Krieg-Brückner: Qualitative Spatial Modelling of Human Route Instructions to Mobile Robots. In: the 3rd International Conference on Advances in Computer-Human Interactions (ACHI), 2010.

3. (45%, 80%, 60%) **Jian**, Zhekova, Shi and Bateman: Deep Reasoning in Clarification Dialogues with Mobile Robots. In: 19th European Conference on Artificial Intelligence (ECAI), 2010.

4. (40%, 80%, 20%) Shi, **Jian** and Rachuy: Evaluation of a Unified Dialogue Model for Human-Computer Interaction. In: International Journal of Computational Linguistics and Applications (IJCLA), 2011.

5. (30%, 80%, 70%) **Jian**, Schafmeister, Rachuy, Sasse, Shi, Schmidt and Steinbüchel: Towards Effective, Efficient and Elderly-friendly Multimodal Interaction. In: the 4th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA), 2011.

6. (30%, 80%, 70%) **Jian**, Schafmeister, Rachuy, Sasse, Shi, Schmidt and Steinbüchel: Evaluating A Spoken Language Interface of A Multimodal Interactive Guidance System for Elderly Persons. In: the 5th International Conference on Health Informatics (HealthInf), 2012.

7. (30%, 80%, 70%) **Jian**, Shi, Schafmeister, Rachuy, Sasse, Schmidt, Hoemberg and Steinbüchel: Touch and Speech: Multimodal Interaction for Elderly Persons. In:

Biomedical Engineering Systems and Technolgies 2012, Lecture Notes, Communication in Computer and Information Science (CCIS), 2013.

8. (30%, 80%, 70%) **Jian**, Shi, Schafmeister, Rachuy, Sasse, Schmidt, Steinbüchel: Better Choice? Combining Speech And Touch In Multimodal Interaction For Elderly Persons. In: the 6th International Conference on Health Informatics (HealthInf), 2013.

9. (60%, 100%, 80%) **Jian** and Shi: Resolving Conceptual Mode Confusion with Qualitative Spatial Knowledge in Human-Robot Interaction. In: the 11th International Conference on Spatial Information Theory (COSIT), Lecture Notes in Computer Science (LNCS), 2013.

10. (30%, 80%, 70%) **Jian**, Shi, Schafmeister, Rachuy, Sasse, Schmidt, Hoemberg and Steinbüchel: Modality Preference in Multimodal Interaction for Elderly Persons. In: Biomedical Engineering Systems and Technolgies 2013, Lecture Notes, Communication in Computer and Information Science (CCIS), 2014.

11. (60%, 100%, 80%) **Jian** and Shi: A Conceptual Model for Human-Robot Collaborative Spatial Navigation. In: the 10th Asia-Pacific Conference on Conceptual Modelling (APCCM), 2014.

SimSpace: A Tool to Interpret Route Instructions with Qualitative Spatial Knowledge

Cui Jian, Hui Shi and **Bernd Krieg-Brückner**

Universität Bremen, Germany
{*ken, shi, bkb*}*@informatik.uni-bremen.de*

Abstract

This paper describes our work on using qualitative spatial interpretation and reasoning to achieve a natural and efficient interaction between a human and an intelligent robot on navigation tasks. The *Conceptual Route Graph*, which combines conventional route graphs and qualitative spatial orientation calculi, serves as an internal model of human spatial knowledge on top of the robot's quantitative representation, such that humans' qualitative route instructions can be interpreted according to the model. The tool *SimSpace* then visualizes and proves the interpretation using qualitative spatial reasoning. Furthermore, *SimSpace* will generate appropriate natural feedback if a route instruction cannot be interpreted properly.

Introduction

Since almost every interactive system uses knowledge of a certain domain to communicate with users, the representation of such domain knowledge decides not only the content but also the manner of the interaction. We focus here on conversational communication between a human and an intelligent service robot (e.g. the Bremen Autonomous Wheelchair *Rolland* (Lankenau and Röfer 2001; Mandel, Huebner, and Vierhuff 2005)) on navigation tasks. One typical scenario is that a human instructs Rolland to move around in a university building with a sequence of route instructions such as "turn left", "pass by the room 1.45 on the left". Although most robots use quantitative information for navigation, such quantitative data are often simplified or even distorted by humans; instead, qualitative representation is often used for representing humans' spatial knowledge and for reasoning with and about it (cf. (Michon and Denis 2001; Shi and Tenbrink 2005)).

Considering this incompatibility of spatial representations between humans and robots, we have developed a qualitative spatial model, i.e., the Conceptual Route Graph, which serves as an internal model for natural communication with humans and for efficient mapping to the robot's quantitative representation. For the automatic interpretation of and reasoning about humans' route instructions, the tool SimSpace has been developed, in which qualitative spatial reasoning is used.

Conceptual Route Graph

Route graphs have been proposed as a common knowledge base of humans or mobile agents for navigation (Werner, Krieg-Brückner, and Herrmann 2000). They are constructed through the integration of a number of routes between different places, where the information concerning accessibility of each place is also integrated. Thus, they can be used as metrical maps to control the navigation of mobile robots in various environments. On the other hand they are used to represent humans' topological knowledge on the qualitative level while they act in space.

The Double-Cross Calculus (DCC) was introduced in (Freksa 1992) for qualitative spatial representation and reasoning using orientation information. Combining the front/back and the left/right dichotomy, the DCC may distinguish 15 meaningful qualitative orientation relations (or DCC relations), such as "front", "rightfront", "right", etc.

The Conceptual Route Graph (CRG) (Krieg-Brückner and Shi 2006; Shi and Krieg-Brückner 2008) combines the structure of conventional route graphs and the Double-Cross Calculus. A CRG is a special graph, its nodes are called *places* and edges *route segments*. Each place has a local orientation, which may be rooted in a global reference frame. Additionally, it has a set of DDC relations describing the orientation relations between route segments and places. A *Route* of a CRG is then a sequence of connected route segments. Thus, CRGs can be seen as route graphs with only qualitative information, i.e. the DCC relations.

SimSpace

SimSpace is a tool for interpreting, visualizing and proving of natural route instructions using qualitative spatial reasoning with a given conceptual route graph. The following are its two most essential functions:

- **Construction of CRGs**: One possible way of constructing a CRG is based on quantitative spatial data. SimSpace takes a well-defined quantitative route graph as input and constructs a corresponding CRG in two steps:
  - the qualification of the quantitative data with the qualifying module of the toolbox SparQ (Wallgrün et al.

2007);

– the generalization of the qualified relations, i.e., the relations qualified from angles near $0°$, $90°$ and $180°$ are assigned to those exactly from $0°$, $90°$ and $180°$, e.g., "rightfront" to "front", or "leftback" to "back". The generalization is necessary for three reasons: first, CRGs serve as an internal model of humans' spatial knowledge and humans tend to use abstracted information while they act in space (cf. (Sadalla and Montello 1989; Montello 1991)); second, ungeneralized relations could be too complicated for qualitative spatial reasoning; third, in most office building environments corridors are constructed orthogonally, thus such generalization retains the environment information.

- **Reasoning with CRGs**: The reasoning with CRGs is based on the following operation:

$$Rel(ab, p) = comp\_path(shortestPath(ab, p)),$$

which calculates the orientation relation between a place p and a segment ab through sequential compositions along the shortest path from ab to p. For example, if the shortest path from x_1x_2 to x_4 is $\{x_1, x_2, x_3, x_4\}$, the relation between x_4 and x_1x_2 can be obtained by:

$$\begin{aligned} Rel(x_1x_2, x_4) &= comp\_path(\{x_1, x_2, x_3, x_4\}) \\ &= comp(Rel(x_1x_2, x_3), Rel(x_2x_3, x_4)) \end{aligned}$$

Using this basic operation, other high-level operations concerning specific route instructions can be defined.

Together with the calculation module provided by SparQ and an ontology-based annotated database, SimSpace supports now a number of often used route instructions, such as "drive straight", "turn", "drive until", "pass by", etc. Through simple actions like selecting and clicking, the interpretation results of given route instructions, i.e., the planning of relevant routes or meaningful feedbacks concerning the spatial mismatches detected in the route instructions, can be proved and generated by SimSpace, respectively. For instance, the route instruction "pass by room A and then room B", which is known as difficult to solve with quantitative spatial computation, can be treated by SimSpace, and the following feedback will be generated, if room B is located before room A from the point of view of the start position:

"Cannot pass by room B, maybe it's now behind you?"

Conclusion

In this paper we presented our work on the modelling and reasoning of humans' natural route instructions using the qualitative spatial model Conceptual Route Graph and the qualitative reasoner SparQ. After building the qualitative spatial model from a given quantitative one, the tool SimSpace provides a set of functions to interpret and prove route instructions according to the qualitative model, and to generate clarification subdialogues in the case of inconsistency of a route instruction with respect to the model. Thus, with SimSpace it is possible to decide whether a spoken route instruction is interpretable by a qualitative spatial model. Consequently, some interpretation(s) will be presented, or adequate reasons will be generated for the further communication with the user.

A large number of route instructions given to the wheelchair Rolland in a university office building was collected in an empirical study (Shi and Tenbrink 2005). We are now using the tool SimSpace to evaluate the coverage of our conceptual model for interpreting those route instructions and to analyze the reasoning results with SparQ for reporting inconsistent situations intuitively.

References

Freksa, C. 1992. Using Orientation Information for Qualitative Spatial Reasoning. In *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, volume 639 of *Lecture Notes in Computer Science*, 162–178. Springer-Verlag.

Krieg-Brückner, B., and Shi, H. 2006. Orientation Calculi and Route Graphs: Towards Semantic Representations for Route Descriptions. In *Proceedings of GIScience 2006, Münster, Germany*, volume 4197 of *Lecture Notes in Computer Science*.

Lankenau, A., and Röfer, T. 2001. A Safe and Versatile Mobility Assistant. *IEEE Robotics and Automation Magazine* 7:29–37.

Mandel, C.; Huebner, K.; and Vierhuff, T. 2005. Towards an Autonomous Wheelchair: Cognitive Aspects in Service Robotics. In *Proceedings of Towards Autonomous Robotic Systems(TAROS 2005)*, 165–172.

Michon, P. E., and Denis, M. 2001. When and Why Are Visual Landmarks Used in Giving Directions? In Montello, D., ed., *Spatial Information Theory*. Springer-Verlag. 292–305.

Montello, D. R. 1991. Spatial Orientation and the Angularity of Urban Routes — A Field Study. *Environment and Behavior* 23(1):47–69.

Sadalla, E. K., and Montello, D. R. 1989. Remembering changes in direction. *Environment and Behavior* 21:346–363.

Shi, H., and Krieg-Brückner, B. 2008. Modelling Route Instructions for Robust Human-Robot Interaction on Navigation Tasks. *International Journal of Software and Informatics* 2(1):33–60.

Shi, H., and Tenbrink, T. 2005. Telling Rolland Where to Go: HRI Dialogues on Route Navigation. In *In WoSLaD Workshop on Spatial Language and Dialogue, Delmenhorst, Germany*.

Wallgrün, J. O.; Frommberger, L.; Wolter, D.; Dylla, F.; and Freksa, C. 2007. SparQ: A Toolbox for Qualitative Spatial Representation and Reasoning. In Barkowsky, T.; Knauff, M.; Ligozat, G.; and Montello, D., eds., *Spatial Cognition V: Reasoning, Action, Interaction: International Conference Spatial Cognition 2006*, volume 4387 of *Lecture Notes in Computer Science*, 39–58.

Werner, S.; Krieg-Brückner, B.; and Herrmann, T. 2000. Modelling Navigational Knowledge by Route Graphs. In C.Freksa; Habel, C.; and Wender, K., eds., *Spatial Cognition II*, volume 1849 of *Lecture Notes in Artificial Intelligence*, 295–317. Springer-Verlag.

Qualitative Spatial Modelling of Human Route Instructions to Mobile Robots

Hui Shi Cui Jian Bernd Krieg-Brückner
SFB/TR8 Spatial Cognition, Universität Bremen, Germany
DFKI Safe and Secure Cognitive Systems, Bremen, Germany
{*shi,ken,bkb*}*@informatik.uni-bremen.de*

***Abstract*—This paper describes our work on the detection and clarification of spatial mismatches in human route instructions while they interact with mobile robots on navigation tasks. To represent and reason about spatial relations in route instructions, a user-centered qualitative spatial model – the Conceptual Route Graph – is introduced. Three reasoning strategies based on this conceptual model are discussed, with which different clarification responses can be generated. Moreover, results of an empirical study to evaluate and compare their effects on people's navigation activities are presented.**

***Keywords*-qualitative spatial representation and reasoning; human-robot interaction; mode confusion; user focus;**

I. Introduction

Imagine that you instruct a mobile robot to navigate in a partially known environment: you are likely to make some *knowledge-based mistakes*, since describing a route is a high-level cognitive process and involves the assessment of complex environment information, such as different spatial frames of reference, localization of spatial objects, and spatial relations between these objects [17]. For example, you might instruct the robot to pass a landmark on the left, although it can only be passed on the right in that situation. Then it naturally becomes a crucial question, how the robot is able to detect such a mistake and inform you about the situation, in particular for the navigation space that consists of multiple interacting representations, both qualitative and quantitative [13].

Usually, a global quantitative map is used for the robot to carry out spatial activities (e.g., moving from one place to another, or reorientating itself). However, using a quantitative representation with numerous metrical data to represent users' intentions, reason about users' spatial knowledge and interacting with them naturally is in general very difficult, less efficient, sometimes even impossible. On the other hand, although qualitative spatial calculi and models (cf. [1], [8], [27], [7]) have been used as the mechanisms for representing and reasoning about spatial relations, using a qualitative spatial model as an intermediate level to represent spatial knowledge, to check its consistency, and to clarify possible inconsistent situations has rarely been investigated.

Nowadays, more and more intelligent robots (e.g. service robots) can be found in daily living. Such robots are typical shared-control systems which are collaboratively controlled by an automation system and a human user. Mode confusions, also called *automation surprises*, have been emphatically studied in safety-critical shared-control systems (e.g. autopilots, cf. [19], [10], [3]). A mode confusion occurs when the observed system state is different from the user's mental state, and may occur frequently, especially for the users of modern service robots, who are usually untrained persons or elderly people. To enable these users to communicate with a robot naturally and intuitively, interaction via natural language or gesture becomes more and more important in service robot research (cf. [2], [16]). Interpreting user intentions using such higher-level modalities leads to a new type of mode confusions, called *conceptual mode confusions*. Knowledge-based mistakes such as those introduced above are typical conceptual mode confusions.

This paper focuses on the detection and clarification of conceptual mode-confusions occurring during human-robot interaction on navigation tasks. The qualitative spatial model – Conceptual Route Graph – (cf. [20]) is used for representing and reasoning about users' navigation knowledge. Based on this model three reasoning strategies are developed, and a corresponding evaluation of these reasoning-based clarifications assisting human users to complement their spatial knowledge and to correct possible mistakes is conducted, which provides several relevant results.

This paper is structured as follows: Section II introduces the qualitative spatial model: the *Conceptual Route Graph*. In Section III we present three different reasoning strategies based on the Conceptual Route Graph. The evaluation study is described in Section IV, and some evaluation results are presented in Section V. We discuss our approach and results in Section VI, before concluding in Section VII.

II. Introduction to the Conceptual Route Graph

A number of research initiatives on communication between humans and intelligent robots using natural language have been reported, several studies are concerned with navigation tasks (e.g., [14], [4], [22]). Motivated by the empirical evidence of these studies, a qualitative spatial model for representing and reasoning about people's spatial knowledge has been developed in [20], i.e., the Conceptual Route Graph.

A. The Conceptual Route Graph

The *Conceptual Route Graph* (or CRG) combines the structure of conventional route graphs and qualitative spatial calculi such as the Double-Cross Calculus.

Route Graphs ([25], [12]) have been proposed as a common knowledge base of humans or mobile robots for navigation. They are constructed through the integration of a number of routes between different places, where the information concerning accessibility of each place is also integrated. They can be used as metrical maps to control the navigation of mobile robots in various environments, and they are also a typical model for representing humans' topological knowledge while they act in space.

The Double-Cross Calculus (or DCC) was introduced in [9], [27] for qualitative spatial representation and reasoning using orientation information. Combining the front/back and the left/right dichotomy, the DCC may distinguish 15 meaningful qualitative orientation relations, such as "front", "right front", "right", etc.

A CRG is a special graph, its nodes are called *places* and edges *route segments*. Each place has a local orientation, which may be rooted in a global reference frame. Additionally, it has a set of DCC relations describing the orientation relations between route segments and places (and landmarks). A *Route* of a CRG is a sequence of connected route segments. Thus, CRGs can be seen as route graphs with only qualitative information, i.e. the DCC relations.

B. Interpretation of Route Instructions with CRG

The most frequently used human route instructions to a mobile robot, reported in the studies [14], [4], [22], are *reorientation*, *directed motion*, *passing*, *moving through*, and *moving to*; the CRG enables an intuitive interpretation of such route instructions. Here we introduce the first three types to demonstrate their interpretation using the CRG. Suppose that the robot is currently at the place p_0, facing the place p_1, i.e. its current position is the route segment from p_0 to p_1, represented as $\overline{p_0p_1}$.

Reorientation is typically expressed by directional instructions such as "turn left/right" or "turn around", which may change the orientation of the robot at the current position. The condition for the robot to perform the turn action is that there exists a place p such that the orientation of p with respect to the current position $\overline{p_0p_1}$ is the given direction d, represented as $\langle\overline{p_0p_1}, d, p\rangle$. The robot's new position then becomes $\overline{p_0p}$.

Directed motion usually contains a turn action and a motion action. After a directed motion action, both the place and the orientation of the robot are changed. For example, to interpret the route instruction "take the next corridor on the left", the most important step is to find the first corridor on the left from the robot's current position. Suppose p_2 is the first junction with a left branching, and the first place along this left branching is p, i.e., the orientation of p with respect to $\overline{p_1p_2}$ is *left*, then the robot's new position is $\overline{p_2p}$ after the specific directed motion.

Passing refers to route instructions containing spatial descriptions external to a path, such as "pass the copy room on the right", or just "pass the copy room" without direction information. Suppose p is the place of the landmark to be passed by and p_2 is a place in front of the current position, i.e. $\langle\overline{p_0p_1}, \textit{front}, p\rangle$. Then we should prove whether the orientation of p with respect to $\overline{p_1p_2}$ is *left* or *right*, i.e. $\langle\overline{p_1p_2}, \textit{left}, p\rangle$ (for pass on the left), $\langle\overline{p_1p_2}, \textit{right}, p\rangle$ (for pass on the right), or both (if no orientation information is given).

C. Qualitative Spatial Representation and Reasoning with SimSpace

SimSpace [11] is a tool for interpreting, visualizing and proving of natural route instructions using qualitative spatial reasoning with a relevant CRG. It has two most essential functions: construction of conceptual route graphs, and reasoning with them. Together with the calculation module provided by the spatio-temporal reasoning toolbox *SparQ* [24] and an ontology-based annotated database, SimSpace supports a number of frequently used route instructions, such as "drive straight", "turn", "drive until", or "pass by". The interpretation results of given route instructions, i.e., the planning of relevant routes or meaningful feedbacks concerning spatial mismatches detected in the route instructions, can be proved and generated by *SimSpace*, accordingly.

III. Detecting and Clarifying Conceptual Mode Confusions via Qualitative Spatial Reasoning

As stated in Section I, describing routes to a robot may lead to *conceptual mode confusions*. There are two frequently occurring types of conceptual mode confusions: *Spatial relation mismatches* and *orientation mismatches* (cf. [21], [20]). A *spatial relation mismatch* occurs, if a route description contains incorrect spatial relations between objects in the environment. Take the sample conceptual model of an environment in Fig. 1, where "the copy room" (at place g) and "the mailbox room" (at h) are located left front of the robot whose current position is denoted as $\overline{ab}$. The route description "pass the mailbox room and the copy room" contains a *spatial relation mismatch*, since the mailbox room is located behind the copy room with respect to the robot's egocentric perspective. Furthermore, if a spatial object is orientated incorrectly, an *orientation mismatch* will occur, as in "pass by the copy room on your right" where the copy room can only be passed on the left.

In this section we are going to introduce three reasoning strategies based on the Conceptual Route Graph, to use them to detect conceptual mode confusions in human route instructions, and to show the clarification information given by respective reasoning results.

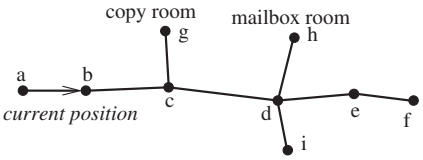


Figure 1. A sample conceptual environment

A. Shallow Reasoning

In order to interpret a route instruction, we should first prove whether the current environment state satisfies the spatial requirements contained in the instruction, such as the possibility to turn left, or to pass a landmark on the right. The *shallow reasoning* strategy is the simplest one of the three reasoning strategies we discuss in this section. It just checks the condition for interpreting a route instruction according to the environment state. If the condition can be satisfied by the environment, the instruction is successfully interpreted, otherwise the route instruction is rejected.

We take the environment in Fig.1 again and interpret the route instruction "pass the copy room on the right". A place has to be found, at which the copy room is located, and whose relation with the current position of the robot $\overline{ab}$ is *right*. Of course such a place cannot be found in this situation; thus, the system rejects the instruction.

B. Deep Reasoning

In contrast, the *deep reasoning* strategy does not simply reject a route instruction, but tries to generate correct spatial relations in contrast to the inconsistent ones contained in the route instruction, i.e., deep reasoning provides relevant spatial information about the specific situation.

Taking the above "pass the copy room on the right" example again, the deep reasoning strategy detects the copy room at the place g and its spatial relation with the robot's current position $\overline{ab}$ as left front, i.e. $\langle \overline{ab}, \mathit{leftFront}, g \rangle$, which is inconsistent with the orientation relation contained in the route instruction. However, the relation $\langle \overline{ab}, \mathit{leftFront}, g \rangle$ describes the correct situation and can be taken as a basis for a clarification dialogue.

The deep reasoning strategy enables the reasoning system to deliver spatial relations that describe specific spatial configurations, e.g., relations between routes and landmarks, which can then be used to explain why a given route instruction cannot be interpreted, assisting people to respond with a better spatial configuration in their navigation instructions.

C. Deep Reasoning with Backtracking

Providing a sequence of route instructions to navigate to a certain place is rather complex: people not only need to locate the current and the destination position correctly – they also have to work in a dynamic mental situation, in which appropriate route instructions need to be constructed and connected with the relevant route, while the imagined current position has to be updated after each instruction is executed mentally. If people make a wrong mental rotation [26] by using a wrong route instruction, and they often do [17], the remaining route instructions will not lead to the desired destination; they might even be uninterpretable since they do not match the subsequent spatial situation.

Consider the route instruction "take the second junction on the right and then drive straight to the mailbox room" in Fig.1. No interpretation can be achieved in this situation, because after the robot turns to the right on the second junction, it cannot find any mailbox room from the current perspective. However, by taking one step backwards, instead of "take the second junction on the *right*", "take the second junction on the *left* and then drive straight to the mailbox room" satisfies the constraints of the situation perfectly.

Therefore, when a route instruction is uninterpretable in a certain situation, instead of simply providing the reason why that instruction is not interpretable, as the deep reasoning strategy does, *deep reasoning with backtracking* tries to find the previously interpreted but potentially incorrect route instruction that caused the failure of the subsequent interpretation, and to provide the possible correction with respect to the problematic instruction, to achieve a successful interpretation of subsequent instructions.

IV. An Empirical Study

In order to explore and compare how people are affected and assisted by the three reasoning strategies based on the Conceptual Route Graph while they are dealing with navigation problems, an empirical study was carried out.

A. Participants

A total of 21 volunteers (scientific researchers and students; 11 female, 10 male) took part in the study.

B. Stimuli and Apparatus

A proper test for the classification of people according to their spatial abilities is itself a research topic (in general, cf. [15], and for navigation in particular, cf. [18]) and not the focus of the current work. A general questionnaire with ten questions describing routes, inquiring for directions, and using maps was used to achieve a coarse classification of the participants with respect to their abilities of mental spatial organization. Because spatial abilities are not uniform across people (cf. [6]), the influences of the three reasoning strategies on people with different spatial abilities should be studied here as well.

Three maps of familiar indoor environments, called the navigation maps, coupled with the three reasoning strategies, were used throughout the study by each participant for the navigation tasks. Every map contained 20 locations (7 named, 13 unnamed) with similar spatial configurations on the same level of difficulty (cf. [5]). The position and

Figure 2. The Chimney House Clinic

orientation of a simulated robot was also given on each map, and remained the same for all participants. In order to help the participants to memorize the maps, all maps were designed with a common object shape layout according to the principle of imagery mnemonics (cf. [23]), cf. "The Chimney House Clinic" in Fig. 2 as an example.

The simulated dialogue system was a networked software application that connected two computers: one computer, called the *navigation assistant*, constantly showed the current system response to the participant, together with a list of the already given route instructions in natural language; the other, called the *brain system*, was responsible for the detection and clarification of conceptual mode confusions, and was only controlled by a human operator who entered the route instructions desired by the participant. The tool *SimSpace* (see Section II-C) is the key component of the brain system. As a result, the whole test run was simulated as if the participant was giving route instructions to the system directly, and getting the feedbacks from the system in natural language as well.

A second questionnaire, called the evaluation questionnaire, concerned the participant's memorization of the map used and his/her feeling about the system responses. The participants completed the questionnaire after finishing the navigation tasks on each map. It was then analysed to examine the impacts of the participant's reasoning strategy on his/her navigation activities.

C. Procedure

The experiment was divided into two phases: learning and testing.

1) Learning Phase: The participant was first asked to fill in the classification questionnaire at the beginning of the learning phase. Then the general information about the test procedure was introduced, including the navigation tasks, the spatial configuration of the maps, and the allowed route instructions (those most frequently used, cf. [22]). In addition, he/she was told to use the navigation assistant and to interpret system responses. A sample map was also presented to the participant, who was then asked to accomplish relevant navigation tasks without looking at the map. Only if the participant mastered the necessary skills, the test would move on to the next phase.

2) Testing Phase: In this phase each participant had to go through three test runs, which corresponded to the three navigation maps coupled with the three reasoning strategies. The sequence of the maps and the combination between each map and each reasoning strategy remained unchanged throughout the study. Each test run consisted of the following three steps:

- *Memorization.* The participant was asked to memorize a map of a familiar indoor environment within exactly one minute; then the map was removed.
- *Navigation.* The participant was free to communicate with the navigation assistant by giving oral route instructions to navigate the mobile robot to three different places on the memorized map. Each task was only finished, when the destination was reached or the participant gave up trying.
- *Evaluation.* At the end of each test run, the participant was asked to fill in the evaluation questionnaire.

V. Evaluation Results

The statistical results focus on the following four criteria which are calculated based on the data recorded throughout the test runs.

- **Satisfaction degree (SD)**: describes how satisfied the participant was by each reasoning strategy. A higher satisfaction degree implies that the participant was better supported by the system responses based on a reasoning strategy. Satisfaction degrees are calculated from the evaluation questionnaire results.
- **Completion degree (CD)**: the degree of successfully completed navigation tasks; the higher the completion degree, the more destinations were reached.
- **Error rate (ER)**: measures the frequency of conceptual mode confusions occurring in each test run.
- **Recidivism number (RN)**: the number of participants, who had repeated errors by giving the same route instructions.

The general results with respect to the above four criteria are presented in Tab. I; they are independent of the participants' abilities of spatial organization. The shallow reasoning, the deep reasoning and the deep reasoning with backtracking strategies are denoted by *RS1*, *RS2* and *RS3*, respectively. Notice that SD with *RS1* is 58.2%, thus satisfaction with *RS1* is just above average (58.2 out of 100);

	SD	CD	ER	RN
$RS1$	58.2%	66.6%	83(2.0/des)	13(61.9%)
$RS2$	73.0%	79.3%	99(1.98/des)	3(14.3%)
$RS3$	75.3%	95.2%	107(1.8/des)	2(9.5%)

Table I
DATA FROM A GENERAL PERSPECTIVE

		SD	CD	ER	RN
$RS1$	A	61.6%	63.0%	29(1.7/des)	5(55.6%)
	B	53.7%	71.4%	32(2.1/des)	5(71.4%)
	C	58.4%	66.7%	22(2.2/des)	3(60.0%)
$RS2$	A	74.2%	85.2%	38(1.6/des)	1(11.1%)
	B	73.1%	76.2%	**38(2.4/des)**	2(28.6%)
	C	70.4%	73.3%	23(2.1/des)	0(0.0%)
$RS3$	A	80.0%	100.0%	17(0.6/des)	0(0.0%)
	B	**66.0%**	**85.7%**	**62(3.4/des)**	**2(28.6%)**
	C	80.0%	100.0%	28(1.9/des)	0(0.0%)

Table II
DATA WITH RESPECT TO A COARSE CLASSIFICATION

CD with *RS1* is 66.6%, derived from the fact that 42 out of a total $21 \times 3 = 63$ places were reached; ER with *RS1* is 83, indicating that 83 errors were made with $RS1$, thus 2.0 errors per destination on average; RN with *RS1* is 13, showing that 13 out of 21(61.9%) participants had repeated the same errors; and so on.

As stated in Section IV-B, people with different abilities of mental spatial organization may be affected differently by the three reasoning strategies. Therefore, the general evaluation results are refined in Tab. II according to the participants' abilities of mental spatial organization. Participants in group A (9 out of 21) have relatively strong, participants in group B (7 out of 21) ordinary, and in group C (5 out of 21) relatively weak abilities of spatial organization.

VI. Discussion

Although all three reasoning strategies are supported by the Conceptual Route Graph model, the information carried by them (in the form of system responses) is quite different, and their influences on users diverge as well. Generally, the satisfaction degree (SD) and completion degree (CD) increases from the shallow reasoning, via the deep reasoning to the deep reasoning with backtracking strategy, while the error rate (ER) and the recidivism number decreases (see Tab. I). The positive effect of both deep reasoning strategies is apparent: the satisfaction (SD) improvements from *RS1* to *RS2* and *RS3* are 25.4% and 29.4%, respectively; the completion (CD) improvements are 19.0% and 42.9%; and the error repetition reduces from *RS1* to *RS2* and *RS3* by 76, 9% and 84.6%. The changes on the error rate (ER) are marginal, the reduced rates are 1.0% and 10%.

Now we consider the improvements from *RS2* to *RS3*: on SD it is 3.2%; on CD 20.1%; on ER 10.0%; and on RN 50.0%. The improvements on CD, ER and RN are obvious, though the satisfaction degree, a result derived from the questionnaires, changes very little. This means that the extension of the deep reasoning strategy with backtracking has many positive influences on human navigation knowledge, but they do not feel much more assisted.

Specifically, the introduction of deep reasoning and backtracking improves the navigation behavior of users with strong spatial organization abilities from group A on all four criteria (see Tab. II), by SD the improvements from *RS1* to *RS2* and *RS3*, and from *RS2* to *RS3* are 20.5%, 29.9% and 7.8%, respectively; and by CD, 35.2%, 58.7% and 17.4%. This is also the case for users with weak spatial organization abilities. However, users who have normal spatial organization abilities (i.e. from group B) seem to be confused sometimes by the additional information, particularly by the information gained from backtracking (see the numbers marked in bold in Tab. II).

In summary, the evaluation results show that the *deep reasoning and backtracking* strategy significantly improves the clarification information for users, who are thus able to instruct mobile robots more effectively and efficiently. This implies that the extra information obtained by qualitative spatial modeling and reasoning does help people to understand the environment better and to correct their errors in processing spatial relations and orientations.

VII. Conclusion and Future Work

This paper presented a qualitative spatial model, i.e. the Conceptual Route Graphs, for the representation and reasoning about human route instructions to mobile robots. Three reasoning strategies have been developed and introduced through examples. In order to compare these strategies and to make assertions on their influences on human navigation knowledge, an empirical study was carried out. The collected empirical data show that applying additional mechanisms in the reasoning process, in our case the deep reasoning and the deep reasoning with backtracking, can generate useful spatial information to help people understand the spatial environment and to correct errors while they instruct mobile robots in navigation tasks.

The integration of the Conceptual Route Graph with its reasoning mechanisms into a dialogue system developed in our research center is yet ongoing work; we also plan to add more refined spatial calculi. The evaluation of the role of the reasoning strategies in people's dialogic behaviour is also interesting work to be done.

The approach presented in this paper can also be adapted to other spatial tasks such as object localization.

Acknowledgements

We gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG) through the Collaborative Research Center SFB/TR8 - Project I3-[SharC] "Shared

Control via Interaction". We would like to thank John Bateman and Elena Andonova for giving us useful suggestions on the design of the empirical study; and Desislava Zhekova for helping us in the generation of system responses.

REFERENCES

[1] J. F. Allen. Maintaining knowledge about temporal intervals. *CACM*, 26(11):832–843, 1983.

[2] J. C. Augusto and P. McCullagh. Ambient intelligence: Concepts and applications. *Computer Science and Information Systems*, 4(1):228–250, 2007.

[3] J. Bredereke and A. Lankenau. A rigorous view of mode confusion. In *Proc. of Safecomp 2002, 21st Int'l Conf. on Computer Safety, Reliability and Security*, number 2434 in Lecture Notes in Computer Science, pages 19–31. Springer-Verlag, 2002.

[4] G. Bugmann, E. Klein, S. Lauria, and T. Kyriacou. Corpus-based robotics: A route instruction example. In *Proceedings of IAS-8*, 2004.

[5] J. Cambell. *Map Use and Analysis*. Columbus(OH): McGraw Hill, 2001.

[6] J. B. Carroll. *Human Cognitive Abilities: A Survey of Factor - analytic studies*. Cambridge University Press, 1993.

[7] A. G. Cohn, B. Bennett, J. Gooday, and N. M. Gotts. Qualitative spatial representation and reasoning with the Region Connection Calculus. *Geoinformatics*, 1, 1997.

[8] A. U. Frank. Qualitative spatial reasoning with cardinal directions. In *Proceedings of the Seventh Austrian Conference on Artificial Intelligence*. Springer, 1991.

[9] C. Freksa. Using orientation information for qualitative spatial reasoning. In *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, volume 639 of *Lecture Notes in Computer Science*, pages 162–178. Springer-Verlag, 1992.

[10] M. Heymann and A. Degani. Constructing human-computer interaction in aeronautics. In *Proc. of HCI-Aero 2002*, 2002.

[11] C. Jian, H. Shi, and B. Krieg-Brückner. SimSpace: A tool to interpret route instructions with qualitative spatial knowledge. In *AAAI Spring Symposium on Benchmarking of Qualitative Spatial and Temporal Reasoning Systems*, 2009.

[12] B. Krieg-Brückner, U. Frese, K. Lüttich, C. Mandel, T. Mossakowski, and R. J. Ross. Specification of an ontology for route graphs. In C. Freksa, M. Knauff, B. Krieg-Brückner, B. Nebel, and T. Barkowsky, editors, *Proceedings of Spatial Cognition IV, Chiemsee, Germany*, volume 3343 of *Lecture Notes in Artificial Intelligence*, pages 989–995. Springer, 2004.

[13] B. Kuipers. The Spatial Semantic Hierarchy. *Artificial Intelligence*, 119:191–233, 2000.

[14] S. Lauria, T. Kyriacou, G. Bugmann, J. Bos, and E. Klein. Converting natural language route instructions into robot executable procedures. In *Proceedings of the 2002 IEEE International Workshop on Human and Robot Interactive Communication*, pages 223–228, 2002.

[15] M. Linn and A. Petersen. Emergence and characterisation of gender differences in spatial abilities: A meta-analysis. *Child Development*, 56, 1985.

[16] K. S. Mukasa, A. Holzinger, and A. Karshmer. Intelligent user interfaces for ambient assisted living. In *Proceedings of the First International Workshop IUI4AAL 2008*, 2008.

[17] J. Reason. *Human Error*. Cambridge University Press, 1990.

[18] M. Roger, N. Bonnardel, and L. Le Bigot. Spatial cognition in a navigation task: Effects of initial knowledge of an environment and spatial abilities on route description. In *Proceedings of the 14th European Conference on Cognitive Ergonomics*, 2007.

[19] J. Rushby, J. Crow, and E. Palmer. An automated method to detect potential mode confusions. In *Proc. 18th IEEE Digital Avionics Systems Conference*. 1999.

[20] H. Shi and B. Krieg-Brückner. Modelling Route Instructions for Robust Human-Robot Interaction on Navigation Tasks. *International Journal of Software and Informatics*, 2(1):33–60, 2008.

[21] H. Shi, C. Mandel, and R. J. Ross. Interpreting route instructions as qualitative spatial actions. In T. Barkowsky, C. Freksa, M. Knauff, B. Krieg-Brckner, and B. Nebel, editors, *Proceedings of International Conference Spatial Cognition 2006*, volume 4387 of *Lecture Notes in Artificial Intelligence*, 2008.

[22] H. Shi and T. Tenbrink. Telling Rolland where to go: HRI dialogues on route navigation. In K. Conventry, T. Tenbrink, and J. Bateman, editors, *Spatial Language and Dialogue*. Cambridge University Press, 2009.

[23] J. T.E.Richardson. *Cognitive Psychology - A Modular Course - Imagery*, chapter 5, pages 103–136. Psychology Press, Hove, UK, 1999.

[24] J. O. Wallgrün, L. Frommberger, D. Wolter, F. Dylla, and C. Freksa. A toolbox for qualitative spatial representation and reasoning. In T. Barkowsky, M. Knauff, G. Ligozat, and D. Montello, editors, *Spatial Cognition V: Reasoning, Action, Interaction: International Conference Spatial Cognition 2006*, volume 4387 of *Lecture Notes in Computer Science*, pages 39–58, 2007.

[25] S. Werner, B. Krieg-Brückner, and T. Hermann. Modelling navigational knowledge by route graphs. In *Spatial Cognition II: Integrating Abstract Theories, Empirical Studies, Formal Methods, and Pratical Applications*, volume 1849 of *Lecture Notes in Artificial Intelligence*, pages 259–316. Springer-Verlag, 2000.

[26] C. D. Wickens. Frames of reference for navigation. In D. Gopher and A. Koriat, editors, *Attention and Performance XVII: Cognitive Regulation of Performance: Interaction Of Theory and Application*, pages 112–144. MIT Press, Cambridge, MA, 1999.

[27] K. Zimmermann and C. Freksa. Qualitative spatial reasoning using orientation, distance, and path knowledge. *Applied Intelligence*, 6:49–58, 1996.

Deep Reasoning in Clarification Dialogues with Mobile Robots

Cui Jian[1] and **Desislava Zhekova**[2] and **Hui Shi**[3] and **John Bateman**[4]

Abstract. This paper reports our work on qualitative reasoning based clarification dialogues in human-robot interaction on spatial navigation. To interpret humans' route instructions, a qualitative spatial model is introduced which represents the robot's beliefs in the application domain. Based on the qualitative spatial model, three tool-supported reasoning strategies are discussed which enable the robot to generate dialogues with differing degrees of clarification if knowledge mismatches or under-specifications in route instructions are detected. The influence of the reasoning strategies on human-robot dialogues is evaluated with an empirical study.

1 Motivation

While instructing a mobile robot to navigate in a partially known environment, humans are likely to make some *knowledge-based mistakes*, since describing a route is a high-level cognitive process and involves the assessment of complex environment information—such as different spatial frames of reference, localization of spatial objects, and spatial relations between these objects [13].

Several research efforts on natural language communication between humans and intelligent systems have been reported in the literature. Some have focused on issues such as corpora based primitive functions (e.g., [12, 4]), task-oriented slot-filling based planning assistants (e.g., [2]) or task tree based tutoring systems (e.g., [6]), while others have applied logical formalization and reasoning in dialogue systems (e.g., [16, 9]). In contrast to these directions, we focus here on the spatial application domain and address the particular challenge of developing a conceptual knowledge model that is capable both of maintaining the spatial knowledge of the environment and of providing a semantic framework for interpreting and reasoning about spatial information to support human-robot dialogues.

Conceptual knowledge representations are already used in some systems to enable human-robot interaction. For example, Zender *et al.* [22] present an approach to create conceptual spatial representations of indoor environments with different layers of map abstractions to support human-robot dialogues. Here we apply instead mathematically well-founded qualitative spatial calculi and models (cf. [1, 10, 23, 7]) since these provide not only the semantic interpretation of humans' spatial knowledge but also mechanisms for reasoning with this knowledge. Although qualitative spatial models have been used for representing and reasoning about spatial relations before, their application as a belief model to assist a robot to communicate naturally with a human has received little attention.

[1] University of Bremen, Germany, email:ken@informatik.uni-bremen.de
[2] University of Bremen, Germany, email:zhekova@uni-bremen.de
[3] DFKI Bremen, Germany, email:hui.shi@dfki.de
[4] University of Bremen, Germany, email:bateman@uni-bremen.de

This paper therefore focuses on the development of a robot's spatial belief model taking qualitative spatial calculi as its foundation. We discuss various tool-supported reasoning strategies which enable the robot to validate route instructions according to its spatial knowledge and create spatial relations that indicate the mismatches or under-specifications in those instructions. Clarification dialogues are then generated according to the relations created. We explore how these more informative clarification dialogues can be provided using qualitative reasoning and evaluate their contribution to effective human-robot interaction systems.

The paper is structured as follows: Section 2 defines the qualitative spatial beliefs model, QSBM, and a set of update rules for interpreting common route instructions based on QSBM. Section 3 introduces three different reasoning strategies and their implementation in the system *SimSpace*. Section 4 then describes an evaluation study and several evaluation results are presented in Section 5. We discuss our approach and results more generally in Section 6, before concluding in Section 7.

2 Modelling a Robot's Spatial Beliefs

Unlike dialogue history or general conversational information in dialogue systems, domain knowledge is application dependent; this requires that a dialogue system be able to talk with the user about domain specific issues. These issues therefore require appropriate representation in their own right. Modern intelligent mobile robots can navigate autonomously in various environments using sensor techniques; but for people to communicate with such robots for joint spatial tasks, an intermediate knowledge representation is necessary so that they can understand each other. The qualitative spatial beliefs model presented in this section is a representation for facilitating such communication.

2.1 Route Graph and Double-Cross Calculus

One proposal for a common knowledge base for humans or mobile agents involved in navigation is that of *route graphs* [20]. Such graphs capture humans' topological knowledge on the qualitative level while acting in space. Route graphs are a special class of graphs. A *node* of a route graph, called a *place*, has a particular position and its own "reference system" for defining directions at that place. An *edge* of a route graph, called a *route segment*, is directed from a source place to a target place and always has three attributes: an *entry*, a *course* and an *exit*.

The Double-Cross calculus (DCC) was introduced by Freksa [11, 23] for qualitative spatial representation and reasoning using orientation information. Combining the front/back and the left/right

dichotomy, the DCC distinguishes 15 meaningful qualitative orientation relations (or DCC relations), such as "front", "right front", "right", etc.

The Conceptual Route Graph (CRG) of Shi and Krieg-Brückner [14] then combines the structure of conventional route graphs and the Double-Cross calculus. The entry and exit of a route segment is further defined by an orientation between the route segment and the reference frame at its source or target place. Additionally, a set of DCC relations are used to describe the orientation relations between route segments and places. A *Route* of a CRG is then a sequence of connected route segments. Thus, CRGs can be seen as route graphs with only qualitative information, i.e. in the present case, the DCC relations.

2.2 QSBM: Qualitative Spatial Beliefs Model

We define a QSBM as a pair of a conceptual route graph and a route segment representing the current position of the robot. We denote this by $\langle crg, pos\rangle$, where the robot is currently located at the entry place of the route segment *pos* and oriented in the direction of its exit place. A conceptual route graph is simply represented by a tuple of four elements $(\mathcal{M}, \mathcal{P}, \mathcal{V}, \mathcal{R})$, where $\mathcal{M}$ is a set of landmarks in the environment, $\mathcal{P}$ a set of topological places, $\mathcal{V}$ a set of vectors from a source place to a target place, and $\mathcal{R}$ a set of orientation relations.

Let p be a place and $\overline{xy}$ a vector from place x to place y. The typical orientations of p with respect to $\overline{xy}$ are then: p is right of $\overline{xy}$ (written as $\langle \overline{xy}, \mathit{Right}, p\rangle$), p is on $\overline{xy}$ (written as $\langle \overline{xy}, \mathit{On}, p\rangle$), or p is in front of $\overline{xy}$ (written as $\langle \overline{xy}, \mathit{Front}, p\rangle$), etc. according to the 15 orientation fields distinguished by the Double-Cross calculus.

In the following discussion we focus on our application scenario, in which we presume that the robot knows essential spatial arrangements (i.e., significant places, connections and orientation relations) of the environment represented by the QSBM.

2.3 Update Rules

Various empirical studies (cf. [12, 4, 15]) confirm that route instructions given by a human to a mobile robot resemble how people instruct other people to navigate (cf. [18, 8]). A common human route description accordingly consists of a sequence of route instructions concerning egocentric or landmark-based reorientations, according to which the QSBM is then progressively updated. In this section we will discuss QSBM update rules for three classes of route instructions used as examples throughout this paper. Each rule has a name, a set of pre-conditions and an effect part. The symbols and operators used in the definitions of update rules are explained where they are first used; in addition we employ the standard operators $\exists$ for the existential quantifier, $\nexists$ for its negation, and $\in$ for the element test operator on sets.

Reorientation is typically expressed by directional instructions like "turn left/right" and "turn around", which may change the orientation of the robot at the current position. The pre-condition for applying this rule is that the robot should find a place in its belief model that satisfies the desired relation and the effect is that it faces that place after the turn operation. Suppose the robot is currently at the place p_0 and faces the place p_1, then the pre-condition is stated as: there exists a place p such that the orientation of p with respect to the route segment $\overline{p_0p_1}$ (or the current position) is the given direction d to turn, i.e., $\langle \overline{p_0p_1}, d, p\rangle$. Formally, we use the rule *Reorientation* to specify the turn-operation, and assume that $\langle(\mathcal{M}, \mathcal{P}, \mathcal{V}, \mathcal{R}), pos\rangle$ is the belief model of the robot with the current position *pos*.

RULE: **Reorientation**
PRE: pos = $\overline{p_0p_1}$, $\exists p \in \mathcal{P}.\langle \overline{p_0p_1}, d, p\rangle$
EFF pos = $\overline{p_0p}$

Moving through instructions, such as "go through the door", usually contain a landmark which should be in front of the robot before, and behind the robot after, the move action. Thus, an important precondition here is to find a route segment (say $\overline{p_2p_3}$) in front of the current position $\overline{p_0p_1}$, such that the landmark is on $\overline{p_1p_2}$. $\overline{p_2p_3}$ is then the new robot position, as specified in the rule *MoveThrough*.

RULE: **MoveThrough**
PRE: pos = $\overline{p_0p_1}$,
$\exists p \in \mathcal{P}, \overline{p_2p_3} \in \mathcal{V}.\mathit{at}(l, p)$
$\wedge\langle \overline{p_1p_2}, \mathit{On}, p\rangle \wedge \langle \overline{p_0p_1}, \mathit{Front}, p_2\rangle \wedge \langle \overline{p_1p_2}, \mathit{Front}, p_3\rangle$
EFF: pos = $\overline{p_2p_3}$

In the rule above l is the landmark referred to in the instruction. The logical conjunction-operator $\wedge$ is used to define the second precondition. Moreover, the relation at associates a landmark with its location.

Passing classifies route instructions containing path external spatial descriptions, such as "pass the copy room on the right" or just "pass the copy room" without direction information. Again, as a precondition the robot should identify the landmark given in the instruction and check whether there is a route such that the landmark is before it at the beginning and behind it at the end of the motion (see the second pre-condition in the following rule). If the direction of passing the landmark is presented in the instruction, certain orientation relations should be satisfied as well. The rule *PassRight* specifies the update strategy for the case in which a landmark should be passed on the right.

RULE: **PassRight**
PRE: pos = $\overline{p_0p_1}$,
$\exists p \in \mathcal{P}, \overline{p_2p_3} \in \mathcal{V}.\mathit{at}(l, p)$
$\wedge\langle \overline{p_0p_1}, \mathit{RightFront}, p\rangle \wedge \langle \overline{p_2p_3}, \mathit{RightBack}, p\rangle$
$\wedge\langle \overline{p_0p_1}, \mathit{Front}, p_2\rangle \wedge \langle \overline{p_1p_2}, \mathit{Front}, p_3\rangle$
EFF: pos = $\overline{p_2p_3}$,
if $\nexists \overline{p_4p_2} \in \mathcal{V}.\langle \overline{p_4p_2}, \mathit{RightBack}, p\rangle \wedge \langle \overline{p_0p_1}, \mathit{Front}, p_4\rangle$

Consequently, the current position of the robot will be updated using the shortest route that satisfies the condition.

3 Qualitative Spatial Reasoning about Route Instructions

As stated in Section 1, humans' route instructions may contain knowledge-based mistakes or incomplete information. Take the sample instance of an environment in Fig. 1, where "the copy room" (at place g) and "the mailbox room" (at h) are located left front of the robot whose current position is denoted as $\overline{ab}$. The route instruction "pass by the copy room on the right" causes an orientation mismatch, while "pass the copy room" contains no information about the spatial orientation of the copy room. In this section we introduce three reasoning strategies, based on QSBM, to perform high level interpretation of such route instructions with respect to the introduced update rules (cf. Section 2.3). We also discuss their corresponding implementation in the system *SimSpace*.

3.1 Three Reasoning Strategies

In order to perform the interpretation of a sequence of route instructions, the environment state is checked and then updated based on QSBM. According to how such processes are executed, three reasoning strategies are distinguished.

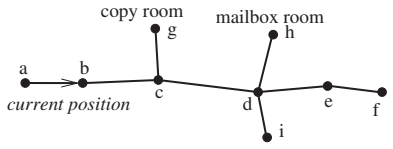


Figure 1. A sample conceptual environment

3.1.1 Shallow Reasoning

The *shallow reasoning* strategy is the simplest one of the three reasoning strategies we discuss. It just checks the condition for interpreting a route instruction according to the environment state. If the condition can be satisfied by the environment, the instruction is successfully interpreted, otherwise the route instruction is rejected.

We take the environment in Fig. 1 again and interpret the route instruction "pass the copy room on the right". A place has to be found at which the copy room is located, and whose relation with the current position of the robot $\overline{ab}$ is *Right*. Such a place cannot be found in this situation and so the system rejects the instruction.

3.1.2 Deep Reasoning

In contrast, the *deep reasoning* strategy does not simply reject a route instruction, but tries to generate correct spatial relations in contrast to the inconsistent ones contained in the route instruction, i.e., deep reasoning provides relevant spatial information about the specific situation.

Taking the above "pass the copy room on the right" example again, the deep reasoning strategy detects the copy room at the place g and its spatial relation with the robot's current position $\overline{ab}$ as left front, which is known from the DCC definitions to be inconsistent with the orientation relation contained in the route instruction.

The deep reasoning strategy then enables the reasoning system to deliver spatial relations that describe specific spatial configurations, and which can be used to explain why a given route instruction cannot be interpreted, assisting people to respond with a more appropriate spatial configuration.

3.1.3 Deep Reasoning with Backtracking

Providing a sequence of route instructions to navigate a robot to a certain place is complex: people not only need to locate the current and the destination position correctly – they also have to work in a dynamic mental situation, in which appropriate route instructions need to be constructed and connected with the relevant route and the imagined current position has to be updated after each instruction is executed mentally. If people make a wrong mental rotation [21] by using a wrong route instruction, and they often do [13], the remaining route instructions will not lead to the desired destination; they might even be uninterpretable since they do not match the subsequent spatial situation.

Consider the route instruction "take the first junction on the left and then drive straight to the mailbox room" in Fig. 1. No interpretation can be achieved in this situation because, after the robot takes the first junction to the left, it cannot find any mailbox room from the current perspective. However, by taking one step backwards, instead of "take the *first* junction on the left", "take the *second* junction on the left and then drive straight to the mailbox room" satisfies the constraints of the situation perfectly.

Therefore, when a route instruction is uninterpretable in a certain situation, instead of simply providing a reason as in *deep reasoning*, our third strategy, *deep reasoning with backtracking*, tries to locate this potentially incorrect instruction, runs the corresponding forward checking of the subsequent route instructions with all possible corrections, and finally achieves a successful interpretation if it exists.

3.2 SimSpace: An Implementation

The qualitative spatial belief model QSBM and its accompanying reasoning strategies has been implemented in the tool *SimSpace*. This is now able to interpret most commonly used route instructions including those discussed in this paper, and to generate *suggest* or *inform* responses if knowledge mismatches or underspecifications are detected.

Given a route instruction parsed in a pre-defined semantic form, *SimSpace* interprets it by automatically selecting an applicable update rule and instantiating its pre-conditions. Taking the sample route instruction "pass the copy room on the right" as an example, *SimSpace* selects the update rule *PassRight* to interpret it. Assuming $\overline{ab}$ to be the current position, the second pre-condition is instantiated to

$$\exists p \in \mathcal{P}, \overline{p_2 p_3} \in \mathcal{V}.at(\text{CopyRoom}, p) \\ \wedge \langle \overline{ab}, \mathit{RightFront}, p\rangle \wedge \langle \overline{p_2 p_3}, \mathit{RightBack}, p\rangle$$

Our adoption of a standard qualitative calculus then allows us to employ newly emerging generic tools for qualitative reasoning. Thus, using the qualitative spatial reasoner *SparQ* [19], the instantiated pre-conditions are checked against the current state of the QSBM and one of our pre-specified strategies. In the case when the shallow reasoning strategy is used, the instruction will be simply rejected by the robot with the response "Not possible". However, using the deep reasoning strategy to interpret the same instruction, we get the result:

$$at(\text{CopyRoom, g}), \langle \overline{ab}, \mathit{LeftFront}, g\rangle$$

i.e., the copy room is located at g and on the left side of the current position. Thus, the orientation relation $\langle \overline{ab}, \mathit{RightFront}, p\rangle$ in the pre-condition cannot be satisfied. Consequently, *SimSpace* translates these results into a corresponding expression in the sentence planning language *SPL* in order to allow generation of a corresponding natural language response using the *KPML* natural language generator [3]. In this case, the response: "I cannot pass the copy room on the right, but I can pass it on the left." is generated.

Finally, by adding the backtracking strategy, every state of QSBM and its interpreted instruction are recorded in a transition history before interpretation proceeds. If any checking/updating process fails, *SimSpace* reloads the previous states and calculates a possible alternation/correction with respect to the interpreted instruction and, if possible, proceeds with the interpretation of the found alternation/correction. Therefore, the response to the instruction "take the first junction on the left and then drive straight to the mailbox room" is then "if I take the second junction on the left, I can drive straight to the mailbox room".

4 An Empirical Study

In order to explore how human-robot dialogues concerning navigation tasks are influenced by our three reasoning strategies, an empirical study was carried out as follows.

4.1 Participants

6 scientific researchers and 15 students, i.e., a total of 21 participants (11 female, 10 male) took part in the study voluntarily.

4.2 Stimuli and Apparatus

Three maps of common indoor environments, called the navigation maps, were coupled with the three reasoning strategies and used throughout the study by each participant for the navigation tasks. Each map contained 20 locations (7 named, 13 unnamed) and was made with similar spatial configurations with the same level of complexity (cf. [5]); in addition, in order to help the participants to memorize the maps, all maps were designed with a common object shape layout according to the principle of imagery mnemonics (cf. [17]); cf. "The Chimney House Clinic" in Fig. 2. Moreover, to retain the start configuration, the position and orientation of a simulated robot was prescribed on each map and this remained the same for all participants.

Figure 2. The Chimney House Clinic

The simulated dialogue system was a networked software application that connected two computers: one computer, called the *navigation assistant*, constantly displayed the current system response and a list of the given route instructions in natural language to the participants; the other, called the *brain system*, was controlled by a human operator who entered the route instructions given by the participant. The *SimSpace* tool (Section 3.2) is the key component of the brain system, checking the instructions using a pre-specified reasoning strategy according to the robot's QSBM and generating clarification dialogues if necessary. As a result, the whole test run was simulated as if the participant communicates with the system in natural language directly, but removing possible distractors that might have been introduced by speech recognition or parsing errors.

An evaluation questionnaire concerning the participant's memorization of the map used and his/her feeling about the system responses was then completed.

4.3 Procedure

Each test was divided into two phases: learning and testing.

4.3.1 *Learning Phase*

Each participant was given a general introduction to the test procedure, including the navigation tasks, the spatial configuration of the maps, commonly used route instructions (cf. [15]) and the ways to interact with the navigation assistant. A sample map was also presented to the participant, who was then asked to accomplish several prescribed sample navigation tasks. The test would only move on to the next phase when the participant had acquired the information necessary.

4.3.2 *Testing Phase*

In this phase each participant had to go through three test runs, which were coupled with the three navigation maps and the three reasoning strategies. The sequence of the maps and the combination between each map and each reasoning strategy remained unchanged throughout the study. Each test run consisted of the following three steps:

- *Memorization.* The participant was asked to memorize a given map of a common indoor environment within exactly one minute; then the map was taken away.
- *Navigation.* The participant was free to communicate with the navigation assistant by giving oral route instructions to navigate the mobile robot to three different places on the memorized map. Each task was only finished when the destination was reached or the participant gave up trying.
- *Evaluation.* At the end of each test run, the participant was asked to fill in the evaluation questionnaire.

5 Results and Analysis

Regarding the dialogue that was formed in the context of the second step of the tests—the *Navigation* step—several interesting points emerged.

In respect to the first and most simplistic reasoning strategy (shallow reasoning: cf. Section 3.1.1), we observed repetition of the mistakes the participants made in their instructions, confusion, and failure to reach the goal as the three most noticeable problems. The repetitiveness of mistakes was mainly caused by the uninformative nature of the system answers (*OK* or *Not possible*). Consider the following dialogue turns:

User: Drive through the door.
System: OK.
User: Turn left.
System: OK.
User: Drive to the end of the corridor.
System: OK.
User: Turn right.
System: OK.
User: Take the second turning on the right.
System: Not possible.

Since from the given dialogue with the system the user can conclude that only the last instruction was wrong, he/she makes an attempt to change only this one. Consequently in the next trial one of the possible replacement utterances could be the following:

User: Take the third turning on the right.
System: Not possible.

The latter is again wrong, but the user does not receive more information about what exactly is wrong with it. Therefore the user keeps

the same course of action and changes again and again the last instruction.

The lack of informativeness in the system answers led as well to another negative result: the confusion and helplessness that accumulated with each negative system answer. Let us consider the following situation:

User: Drive through the door.
System: Not possible.

In this case the user correctly navigated to the dining room. The only mistake was that the robot was not oriented in the direction of the goal (see the update rule *MoveThrough* in Section 2.3). However, receiving an answer *"Not possible"* to the instruction *"Drive through the door"* confuses the user (i.e. what the user might think is that if it is not possible to enter the dining room then the position of that room must be somewhere else). Thus, the navigator tries to change the correctly remembered position of the dining room and find another position for it in the mental representation of the map he/she has. In other words, the user creates a new spatial mismatch as a result of the system answer. The further negative answers of the system, which lack any reasoning about why it is not possible to drive through the door at that point, only make the user give up trying to reach the goal. Consequently only 42 of the 63 goals altogether were reached by the participants with the help of the shallow reasoning strategy.

Only giving instructions and getting *"OK"* or *"Not Possible"* as answers can hardly be considered helpful dialogue. On the other hand, additionally giving a reason for why it is not possible to perform a certain action certainly can. Reasoning does not only improve on the naturalness of the dialogue, but on its usefulness as well. We consider two very simple instructions in the first and second reasoning strategies in Table 1. Naturally, after receiving the answers to the instructions in Strategy #1 (shallow reasoning), the user does not know what went wrong and what instruction would result in a successful completion of the task. In Strategy #2 (deep reasoning) however, the user receives enough information in order to give an instruction that could lead either to the completion of the task or to a correction of the previous false instruction. We consider the latter to be an important factor that participates considerably to the increase in the number of trials in which goals were successfully reached: i.e., 48/63.

Str.	Answer
#1	Not possible.
#2	We can not drive until the lab, because it is behind us.

Table 1. Example system answers from Strategy #1/#2 to the instruction *Drive until the lab.*

A further improvement of the dialogue capabilities is provided in the third reasoning strategy—Strategy #3 (deep reasoning with backtracking); this is the possibility for correcting false instructions by giving a suggestion for an action that is closest to the given command and also possible to perform (see Section 3.1.3). Examples of such dialogue turns are given in Table 2. In natural conversations humans are normally able to give a reason if they can not perform a certain action, but it is not always the case that they can suggest what could be done to make corrections due to the lack of such knowledge. Thus, we consider that the deep reasoning strategy with backtracking moves us closer to a cognitive and helpful natural language human-human dialogue. As shown in Table 3, 58 goals from altogether 63 were found in comparison to the 42 goals out of 63 with the simplest reasoning strategy. This supports our assessment of Strategy #3 as being more helpful to participants.

Str.	Answer
#1	Not possible.
#3	If I take the first junction to the left, I can't drive until the Communication department on the right, but if I take the second junction to the left, I could drive until the Communication department on the right.

Table 2. Example system answer from Strategy #1/#3 to the instruction *Take the first junction to the left and then drive to the Communication Department to the right*.

In the second row of Table 3 are figures representing the average number of instructions in the cases in which the participants reached the goal. The fact that this number is increasing for the second and third reasoning strategy demonstrates that the users actually had longer dialogues with the system to reach the goals and did not rely only on their own memory but on the information provided in the system answers as well.

Last but not least, we can pay attention to the last row of Table 3 where we show the satisfaction degree of the participants indicating how content they were with the system answers for each reasoning strategy. According to these questionnaire results, the shallow parsing strategy is ranked as least sufficient with a result almost 15% below the next reasoning approach. Deep reasoning with backtracking received only about 2% higher result than deep reasoning. This according to our observations is an effect brought about by the increasing complexity of the system responses when backtracking is added. Nevertheless, both latter strategies appeared to be appreciated well by the participants.

The results of this experiment serve to demonstrate the influence of clarification dialogues with mobile robots as well as to point out the direction in which the system should be further developed. They accordingly provide a base for more exhaustive evaluation that has been planned for the further steps of the development.

	Str. #1	Str. #2	Str. #3
Reached Goals	66.67%	76.19%	92.06%
Average Instructions per Reached Goal	10.17	11.15	13.67
Satisfaction Degree	58.2%	73.0%	75.3%

Table 3. Summary of the experimental results.

6 Discussion

Spatial mismatches can often be observed in human route instructions in everyday life, but they usually do not pose a great problem in human-human communication since humans are able to easily spot and clarify the mismatch at hand. In human-robot interactions during navigation tasks, however, such spatial mismatches can cause difficulties and complications as well as predisposing users negatively to the system.

In our work we have shown that deep reasoning with and without backtracking in clarification dialogues with mobile robots can help spot, exemplify and reduce to a great extent the spatial mismatches in human-robot dialogue. Moreover, once the mismatch is identified, an informative and helpful correction and suggestion increases the usefulness of the dialogue itself. This makes the latter more constructive, valuable and helpful—characteristics still problematic in the area of human-robot dialogue systems at large.

Furthermore, we have also treated some new specific challenges for situated dialogue in the current work. The issue is not only that more informative feedback helps users complete a larger number of direction-giving tasks, which is indeed not surprising, but precisely how that needs to be done with what kind of additional information. It is crucial to explore this precise point of contact if effective dialogue systems for human-robot are to be built: we need to know more about just what information is needed where. This is explored directly in the experimental setup pursued.

Thus, such reasoning-based informative dialogues should be considered a highly desirable design feature for situated and general dialogue systems that address the need of improving human-robot dialogue so as to resemble human-human interaction more closely.

7 Conclusion and Future Work

This paper has reported work that integrates several interests of Artificial Intelligence. Concretely we treat the following aspects in depth: the management and formalization of, and reasoning with, domain knowledge. All three represent essential components for human-robot dialogue systems. Specifically, we presented three reasoning strategies and discussed their influences on clarification dialogues with mobile robots. The major contributions of the current work are twofold: the development of the robot's qualitative spatial belief model (QSBM), and the generation of the robot's responses using qualitative reasoning. Furthermore, a preliminary empirical study confirmed that qualitative spatial reasoning mechanisms, in our case the deep reasoning and the deep reasoning with backtracking strategies, provide useful information for the robot to generate more natural and informative dialogues. This therefore encourages further experiments based on these results.

The integration of the qualitative spatial belief model, including the reasoning mechanisms and their implementation in *SimSpace*, into a natural language dialogue system is still ongoing work. We are also planning to add more refined qualitative calculi and reasoning abilities to enrich both the descriptions of the application domains and the sophistication of the system's spatial responses. Moreover, the research on the close interaction between knowledge management and dialogue modelling/control, e.g., the application of our reasoning methods to catch misunderstandings of the system, will also be pursued further.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG) through the Collaborative Research Center SFB/TR 8 Spatial Cognition - Subprojects I3-SharC and I5-DiaSpace, and Department of Safe & Secure Cognitive Systems at German Research Center for Artificial Intelligence (DFKI) Bremen.

REFERENCES

[1] James F. Allen, 'Maintaining knowledge about temporal intervals', *CACM*, **26**(11), 832–843, (1983).

[2] James F. Allen, Lenhart K. Schubert, George Ferguson, Peter Heeman, Chung Hee Hwang, Tsuneaki Kato, Marc Light, Nathaniel G. Martin, Bradford W. Miller, Massimo Poesio, and David R. Traum, 'The trains project: A case study in building a conversational planning agent', *Journal of Experimental and Theoretical AI*, **7**, 7–48, (1994).

[3] John A. Bateman, 'Enabling technology for multilingual natural language generation: the KPML development environment', *Journal of Natural Language Engineering*, **3**(1), 15–55, (1997).

[4] Guido Bugmann, Ewen Klein, Stanislao Lauria, and Theocharis Kyriacou, 'Corpus-based robotics: A route instruction example', in *Proceedings of IAS-8*, (2004).

[5] John Cambell, *Map Use and Analysis*, Columbus(OH): McGraw Hill, 2001.

[6] Brady Clark, Oliver Lemon, Alexander Gruenstein, Elizabeth Owen Bratt, John Fry, Stanley Peters, Heather Pon-Barry, Karl Schultz, Zack Thomsen-Gray, and Pucktada Treeratpituk, *Advances in Natural Multimodal Dialogue Systems*, chapter 13: A General Purpose Architecture for Intelligent Tutoring Systems, 287–305, Springer Netherlands, 2005.

[7] Anthony G. Cohn, Brandon Bennett, John Gooday, and Nicholas Mark Gotts, 'Qualitative spatial representation and reasoning with the region connection calculus', *GeoInformatica*, **1**(3), 275–316, (1997).

[8] Michel Denis, 'The description of routes: A cognitive approach to the production of spatial discourse', *Cahiers de Psychologie Cognitive*, **16**, 409–458, (1997).

[9] Debora Field and Allan Ramsay, 'Deep-reasoning-centered dialogue', in *The 11th European Workshop on Natural Language Generation*, pp. 131–138, Morristown, NJ, USA, (2007). Association for Computational Linguistics.

[10] Andrew U. Frank, 'Qualitative spatial reasoning with cardinal directions', in *Proceedings of the Seventh Austrian Conference on Artificial Intelligence*, Springer, (1991).

[11] Christian Freksa, 'Qualitative spatial reasoning', in *Cognitive and Linguistic Aspects of Geographic Space*, eds., D. M. Mark and A. U. Frank, Kluwer, (1991).

[12] Stanislao Lauria, Theocharis Kyriacou, Guido. Bugmann, Johan Bos, and Ewan Klein, 'Converting natural language route instructions into robot executable procedures', in *Proceedings of the 2002 IEEE International Workshop on Human and Robot Interactive Communication*, pp. 223–228, (2002).

[13] James Reason, *Human Error*, Cambridge University Press, 1990.

[14] Hui Shi and Bernd Krieg-Brückner, 'Modelling Route Instructions for Robust Human-Robot Interaction on Navigation Tasks', *International Journal of Software and Informatics*, **2**(1), 33–60, (2008).

[15] Hui Shi and Thora Tenbrink, 'Telling Rolland where to go: HRI dialogues on route navigation', in *Spatial Language and Dialogue*, eds., K. Conventry, T. Tenbrink, and J. Bateman, Cambridge University Press, (2009).

[16] Ronnie W. Smith and D. Richard Hipp, *Spoken Natural Language Dialog Systems: A Practical Approach*, Oxford University Press, 1994.

[17] John T.E.Richardson, *Cognitive Psychology - A Modular Course - Imagery*, chapter 5, 103–136, Psychology Press, Hove, UK, 1999.

[18] Barbara Tversky and Paul U. Lee, 'How space structures language', in *Spatial Cognition: An interdisciplinary Approach to Representation and Processing of Spatial Knowledge*, eds., C. Freksa, C. Habel, and K.F. Wender, volume 1404 of *Lecture Notes in Artificial Intelligence*, pp. 157–175. Springer-Verlag, (1998).

[19] Jan Oliver Wallgrün, Lutz Frommberger, Diedrich Wolter, Frank Dylla, and Christian Freksa, 'A toolbox for qualitative spatial representation and reasoning', in *Spatial Cognition V: Reasoning, Action, Interaction: International Conference Spatial Cognition 2006*, eds., T. Barkowsky, M. Knauff, G. Ligozat, and D. Montello, volume 4387 of *Lecture Notes in Computer Science*, pp. 39–58, (2007).

[20] Steffen Werner, Bernd. Krieg-Brückner, and Theo Hermann, 'Modelling navigational knowledge by route graphs', in *Spatial Cognition II: Integrating Abstract Theories, Empirical Studies, Formal Methods, and Pratical Applications*, volume 1849 of *Lecture Notes in Artificial Intelligence*, pp. 259–316. Springer-Verlag, (2000).

[21] Christopher D. Wickens, 'Frames of reference for navigation', in *Attention and Performance XVII: Cognitive Regulation of Performance: Interaction Of Theory and Application*, eds., Daniel Gopher and Asher Koriat, 112–144, MIT Press, Cambridge, MA, (1999).

[22] Hendrik Zender, O. Martinez Mozos, Patric Jensfelt, G.-J. M. Kruijff, and Wolfram Burgard, 'Conceptual spatial representations for indoor mobile robots', *Robotics and Autonomous Systems*, **56**, (2008).

[23] Kai Zimmermann and Christian Freksa, 'Qualitative spatial reasoning using orientation, distance, and path knowledge', *Applied Intelligence*, **6**, 49–58, (1996).

Evaluation of a Unified Dialogue Model for Human-Computer Interaction

Hui Shi*, Cui Jian, Carsten Rachuy

SFB/TR8 Spatial Cognition, Universität Bremen, Germany
{shi,ken,rachuy}@informatik.uni-bremen.de
*Safe and Secure Cognitive Systems, DFKI Bremen, Germany

Abstract. This paper reports our work on evaluating the task success of a dialogue model developed by a unified dialogue modeling approach for human-computer interaction, which combines an information state based dialogue theory and a state-transition based modeling approach at the illocutionary level. As an application, the unified dialogue model has been integrated into a multimodal interactive guidance system for hospital visitors. An experiment with 12 subjects has been carried out. Using the collected dialogue data we have evaluated the task success of the dialogue model by the Kappa coefficient. The results show that the unified dialogue model is highly effective and provide several valuable improvements for the further development as well.

keywords: human-computer dialogue, dialogue act, illocutionary structure, information state, dialogue system evaluation, formal methods

1 Introduction

Generalized Dialogue Modeling (cf. [14, 8, 12]) and *Information State* based dialogue theories (cf. [15, 5, 2, 4, 7, 16]) are the two most important approaches to develop dialogue models. Generalized dialogue models are based on recursive transition networks. These models consist of pattern-based accounts of dialogue structure at the illocutionary level and therefore, are independent of utterance content or other direct surface indicators. Information state theories, on the other hand, offer a powerful basis for interaction analysis and practical dialogue system construction. However, such information state based dialogue models are difficult to manage, to extend and to reuse. Although it has been suggested that applying generalized dialogue models to information state based accounts could eliminate some of the perceived problems, there have only been preliminary researches to date [18, 8]. In Lewin [8], for example, recursive transition networks were applied to model Conversational Game Theory by combining dialogue grammars with discourse planning.

The unified dialogue modeling approach introduced in this paper combines the information state based dialogue theory discussed in [16, 7] and the generalized dialogue modeling approach proposed in [14, 12]. Specifically, unified

dialogue models extend generalized dialogue models by introducing *context-sensitive* transitions, which allow for direct integration with information state management. A unified dialogue model is represented as the traversal of a state-transition network with arcs denoting context-sensitive transitions and nodes denoting dialogue states. In addition to the allowed dialogue action, each context-sensitive transition is associated with a set of *conditions* under which the dialogue action can be taken and a set of *update rules* for updating the information state after performing the dialogue action.

As emphasized in [11, 12], the separation of illocutionary structures from the information state-based modeling enables the formal analysis and comparison of illocutionary structures by applying well-established techniques from the formal methods community of computer science. In this paper, we focus on the evaluation of unified dialogue models. The *Kappa coefficient* [13, 3] has been proposed as a standard measure of reliability and task success ([17]) for evaluating spoken dialogue systems. Therefore, we apply it to evaluate how well human users can be supported by the unified dialogue model implemented in a multimodal dialogue system for guiding visitors in hospital environments. For this purpose we carried out an experiment with 12 people and collected 272 dialogues.

This paper is organized as follows: Section 2 introduces the unified dialogue modeling approach, which has been applied to develop a unified dialogue model for a practical multimodal dialogue system presented in Section 3. Section 4 describes the experiment and the collected dialogue data, which are then used to evaluate the unified dialogue model by the Kappa coefficient in Sections 5 with respect to the measure of task success. The evaluation results and corresponding improvements are discussed in Section 6. Finally, Section 7 concludes with the outline of future work.

2 A Unified Approach for Dialogue Modeling

The unified modeling approach takes as a starting point existing researches on the generalized dialogue modeling at the illocutionary level using Recursive Transition Networks (RTNs) [14]. Unlike finite state models, the RTNs employed here capture more abstract dialogue models which depict discourse patterns in illocutionary force terms only – without reference to propositional content or other direct surface indicators. Fig. 1(a) depicts a transition diagram named *Assert(A,B)* initiated by a dialogue participant, say A, and responded to by B. The darkened circles denote final states. This generalized transition diagram is initiated by A's dialogue move of type *assert*. The possible responses from B are threefold: B agrees with the assertion (*B.agree*), accepts it (*B.accept*) or rejects it (*B.reject*). To note that, the transition diagrams *Ask(B,A)* and *Assert(B,A)* are used to enable B to ask some question(s) before reacting to A's request, or to give possible reason(s) by a rejection, and are not presented here in detail.

Generalized dialogue models such as the one depicted in Figure 1(a) are non-deterministic models, where more than one dialogue move is able to trigger state transitions starting from one state. The decision as to which transition should

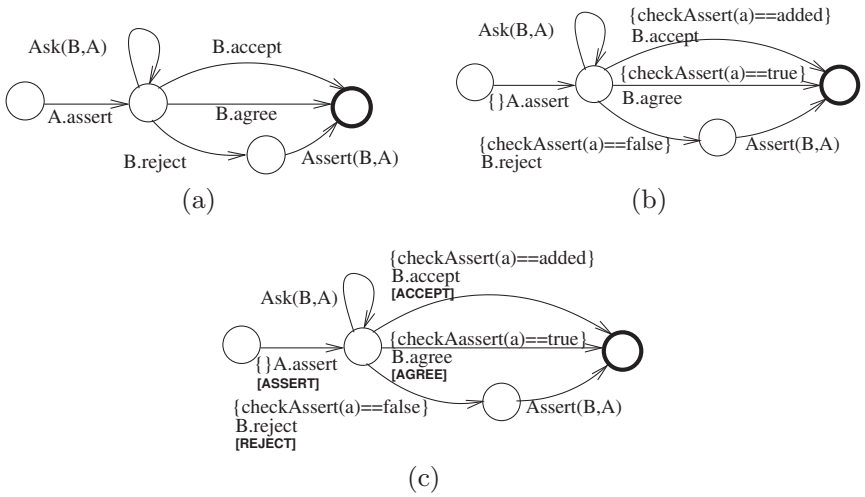


Fig. 1. Three transition diagrams: (a) non-deterministic assertion, (b) deterministic assertion, and (c) deterministic assertion with update rules

be activated naturally depends to a certain extend on B's pragmatic domain knowledge. To take domain knowledge into account, thus to solve such nondeterministic transitions, *conditional transitions* are introduced in unified dialogue models. A conditional transition can be activated only if its conditions are satisfied. Let *checkAssert* be an operation provided by B's domain component, which takes an assertion as a parameter and returns *true* if B's knowledge matches the assertion; or *false* if the assertion conflicts with some of B's knowledge (in that case, the transition diagram *Assert(B,A)* will be activated to explain the reason for B's rejection); or *added*, if the assertion can be added by B as a new element to the knowledge base. A deterministic transition diagram for the example is now shown in Figure 1(b), where a is assumed to be the assertion made by A.

Although conditional transition models as shown in Fig. 1(b) capture the illocutionary structure of dialogues and are deterministic as well, they do not provide mechanisms to integrate dialogue context and history. Therefore, they do not reflect dialogue participants' attitudinal state along with the behavioral mechanisms for dialogue progression and the dynamic update of attitudinal states over time. As indicated earlier, information state based approaches of dialogue models [9, 15, 4] and dialogue management [16, 7] focus on the modeling of dialogue contexts and participants' attitudinal states, apart from that they do not capture the structural features of dialogues. Thus, merging these two approaches is valuable, so that the basic formalism of the conditional transition models is extended by introducing a mechanism to interface with information state.

Since generalized dialogue models already capture structural features of dialogue moves, some of the typical *structural elements* in the information state based accounts, e.g., *AGENDA* for keeping the planed dialogue acts in Ginzburg and Larsson's models, become unnecessary, hence the information model can

be simplified considerably. In unified dialogue models, each transition can be associated with one or more update rules for updating the current information state if needed before proceeding to the next state. As usual, an update rule consists of a name, a set of pre-conditions and a set of operations on information states. To illustrate this model extension, we again take the transition diagram *Assert(A,B)* as an example and show it in Fig. 1(c). After dialogue participant A makes an assertion, the update rule *ASSERT* will be applied to update the information state, such that the new assertion can be integrated into the current information state. Similarly, B's transitions of *accept*, *agree* and *reject* can change the information state by the corresponding update rules.

Finally, a *unified dialogue model* is a pair $\langle \mathcal{G}, G_0 \rangle$ of a transition network $\mathcal{G}$ with a set of extended recursive transition diagrams and a main diagram $G_0 \in \mathcal{G}$. Each transition may contain some conditions and information state update rule(s). Specifically, if a dialogue is in the start state of a transition whose conditions are satisfied, the corresponding dialogue move is then enabled and the information state is updated by its update rules, and the dialogue will move to its goal state.

3 MIGHE: A Multimodel Interactive Guidance for Hospital Environment

MIGHE is a multimodal interaction system developed for guiding people in public areas such as hospitals. Fig. 2 shows the overall MIGHE architecture. This section focuses on the development of a unified dialogue model and its integration into the dialogue system. The unified dialogue model is implemented within the two components: the *dialogue controller* and the *information state manager*. The *clinic database manager* provides the dialogue controller with necessary information about application environment. The dialogue controller manages the communication between various system components, and controls the dialogue process according to the dialogue model together with the information state manager. The guidance system supports both natural language inputs and touch events, but in the experiment presented in Section 4 only the natural language input channel is enabled.

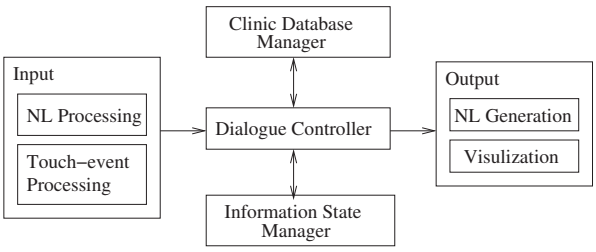


Fig. 2. The overall architecture of the dialogue system

3.1 The Unified Dialogue Model

The dialogue model implemented in MIGHE is developed according to the unified dialogue modeling approach introduced in Section 2. In this paper we focus on the task orientated dialogues and disregard communication problems like failures by speech recognition or misunderstanding. Generally, these problems can be treated by extending the dialogue model. The information state structure consists of two parts: *LM* for keeping the latest dialogue move and *CONTEXT* containing a list of contexts of active (sub-)dialogues. In this application, the possible contexts are of the types: *department*, *person*, or *room*, which provide context information for integrating user's dialogue moves, for example, "go to *a room* of *a known department*", or "request for information of *a person* in *a department*".

The unified dialogue model consists of four extended transition diagrams with the main diagram *Dialogue(S,U)*, see Fig. 3. After a system's initializing *request* (Fig. 3(a)), the user can instruct the system to find some visiting goals by utterances with the dialogue act *instruct*, or ask the system to find certain information by *request* (see *Dialogue(U,S)* in Fig. 3(b)). The network *Response(S,U)* (Fig. 3(c)) specifies all deterministic system responses after getting an input from the user according to its domain knowledge and the current information state. If the requested information or instructed goal does not exist, the user's input is *rejected*, probably with a reason if the relevant information is available. If it is found unambiguously, the user is *informed* and asked whether he/she would like to take the found place as a destination in case the last user input is an instruction. However, if more than one possibility are found, a subdialogue is started by the system for asking the user to make a *choice*. Finally, *Response(U,S)* (Fig. 3(d)) describes possible *nondeterministic* user reactions to a system's request. Moreover, each dialogue move issued by a user in the dialogue model is associated with the name of an update rule.

3.2 Integrating the Unified Dialogue Model into the Dialogue System

The implementation of the unified dialogue model is carried out in two major steps. In the first step, a set of update rules as required by the dialogue model is implemented for the component *information state manager*. Five update rules are needed in the unified dialogue model (see Fig. 3). The following shows the rule INSTRUCT as an example. Suppose that *context* and *dest* are two operations to identify the context and destination contained in an input, respectively. The context and destination of "I'd like to go to Mrs. Angelika Fromm in Gastroenterology", for example, are "Gastroenterology" and "Mrs. Angelika Fromm". If the current instruction contains context information, i.e., the user gives the context in his/her instruction explicitly, then the new context will be added to *CONTEXT*, otherwise, the most actual context in *CONTEXT* (or *top*(*CONTEXT*)) is used to complete the current instruction. The other rules are defined accordingly.

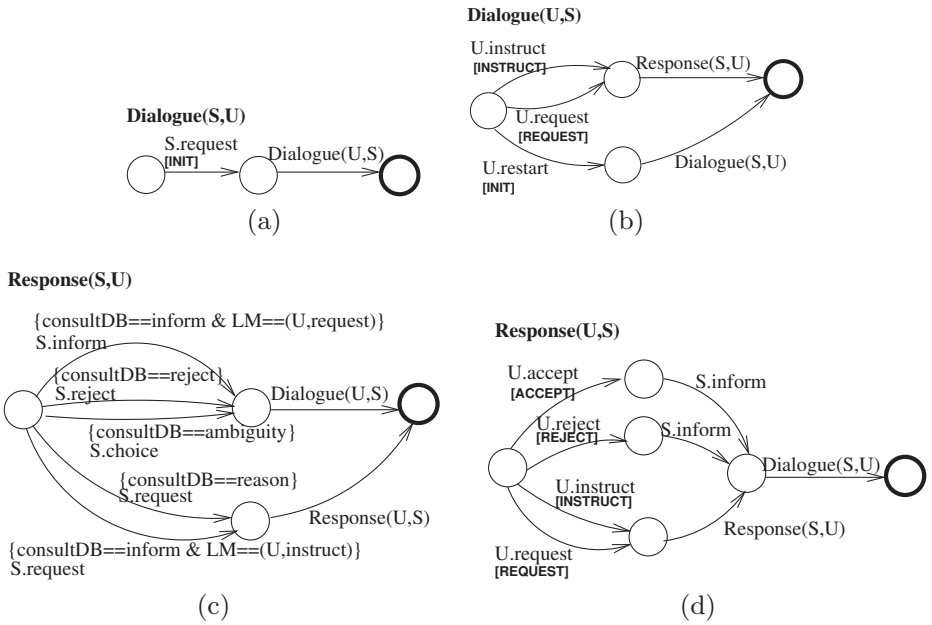


Fig. 3. The unified dialogue model: (a) the main transition diagram, (b) the transitions issued by the user, (c) the system's responses and (d) the user's response

RULE: INSTRUCT
PRE: if $context(m)$!=$null$ then $c = context(m)$ else $c = top(CONTEXT)$, $d = dest(m)$
EFF $LM = (U, instruct)$,
if $context(m)$!=$null$ then $CONTEXT = add(CONTEXT, c)$

The second step is the development of the control mechanism of the component *dialogue controller*, which is based on the dialogue state transitions at the illocutionary level specified by the dialogue model. As the unified dialogue model defines a clear illocutionary structure represented by a set of extended recursive transition diagrams, it can be specified with mathematically well-founded methods straightforwardly, e.g., the well-established technique from the formal methods community of computer science *Communicating Sequential Processes* (CSP). The CSP language provides mechanisms for specifying the communication and synchronization of two or more processes consisting of sequential actions. The essential value of CSP is the ability to subject formal specifications that are well founded in mathematical logic to enable powerful analysis using mechanized theorem provers and model checkers (cf. [12]). Although the CSP language, its mathematical foundations and its many possible applications within the Formal Methods Community have been widely investigated [6, 10], applying these techniques to dialogue modeling, specification and analysis builds up a novel area of application. In the following we will briefly introduce the specification of the unified dialogue model presented in Section 3.1 using CSP.

The first CSP process *DialogueUS* in Fig. 4 specifies the transition network *Dialogue(U,S)*, where → and [ ] are two CSP operators necessary for the present specification. → defines the sequential occurrence of dialogue moves in a process, and [ ] arbitrary selection between several possibilities. The CSP events representing abstract dialogue moves have the form $p.a$, where p is the name of a communication channel and a the dialogue act associated with it. For example, *user.instruct* means getting an input with the dialogue act *instruct* from the user, *is_out.instruct* sending the dialogue act *instruct* to the information state manager, such that the information state can be updated using the context contained in the current input. Obviously, the specification reflects the model structure very well. The second CSP process *ResponseSU* invoked by the first one in Fig. 4 specifies the transition network *Response(S,U)*, in which the *latest dialogue move* kept in the information state is needed. In the specification *ResponseSU* the conditions related to *consultDB* are specified by four database input *db_in* events: *reject*, *reason*, *inform* and *ambiguity*. Also the CSP specification of *Response(U,S)* reflects the network structure straightforwardly.

```
DialogueUS =
    user.restart -> is_out.init -> DialogueSU
 [] user.instruct -> is_out.instruct -> ResponseSU
 [] user.request -> is_out.request -> ResponseSU

ResponseSU = db_out -> (
    db_in.reject -> system.reject -> DialogueUS
 [] db_in.reason -> system.request -> ResponseUS
 [] db_in.inform -> is_in?lm ->
      ( (lm==request) & (system.inform -> DialogueUS)
      [](lm==instruct) & (system.request -> ResponseUS))
 [] db_in.ambiguity -> system.choice -> DialogueUS)
```

Fig. 4. Two CSP specifications

Based on the CSP specifications the model-checker FDR [1] is applied to generate the state machine. After implementing the communication channels between the dialogue controller and the other system components, the state machine can control the state transitions according to communication events.

4 The Experiment

In order to explore how well the dialogue interaction between human and the dialogue system is assisted by the unified dialogue model, an evaluation with 12 participants was carried out. Each subject had to undergo two test phases: learning and testing:

– In the learning phase each participant was given a brief introduction to the test procedure, so that they could get to know the way how to dialogue with the system, and what kinds of verbal and textual feedbacks the system provides. Furthermore, they were asked to accomplish several sample tasks.

– In the test phase each participant had to go through three subphases, each of which contains several tasks belonging to a predefined category. In the first subphase, several pieces of information describing a destination (e.g. a person's name, a department or a room number) were given and the participant should tell the system to go there. In the second subphase, pieces of information were given as well, but this time the participant was asked to find out certain information, e.g. where a certain person works or what department a room is in. In the third subphase scenarios like "you are hungry and would like to eat something" were described, and the participant was asked to negotiate with the system on an appropriate destination.

The dialogue system used in the experiment was a networked software application that connected two computers: the *guidance assistant* on one computer and the *input system* on the other. The *input system* was controlled by a human operator who entered the user utterances and acted as a speech recognizer. The *guidance assistant* contains the components *clinic database manager*, *output*, *information state manager* and *dialogue controller*, and the unified dialogue model is the key of the *information state manager* and *dialogue controller*. As a result, the whole test run was simulated as if the participant communicates with the system in natural language directly, but removing possible distractions that might have been introduced by speech recognition, in order to focus on the evaluation of the unified dialogue model. Although a human operator acted as the speech recognizer, our experiment was not a usual "Wizard of Oz" experiment, since the *guidance assistant* ran automatically.

Since the experiment was run with native German-speaking participants, we present in the following the English translations of several example dialogues collected in the experiment. Most of the dialogues turned out to be unproblematic, as the following example shows. The task of the example contained the destination "Rasmussen" and "Room number 1322". The room number in the user's *instruction* (*D_01_01*) was incorrect, thus the system *rejected* it in *D_01_02* with a reason. In *D_01_03* the user provided the room number again, which is interpreted as a new *instruction* integrated with the context *person* provided in the previous instruction and kept in *CONTEXT* of the information state. This time the system found the person and his office, and then responded with a *request* such that the user can confirm or reject it.

D_01_01 U I'd like to go to Mrs. or Mr. Rasmussen in room number 1320.
D_01_02 S I did find Rasmussen, but not in room 1320.
D_01_03 U I'd like to go to room 1322.
D_01_04 S Would you like to go to Mr. Per Rasmussen in room 1322, the office of chief physician?
D_01_05 U Yes.

However, there are dialogues turned out to be problematic. The utterances from *D_02_01* to *D_02_04* fit well to the model, but the utterance *D_02_05* is an user *request*, for which no proper system response is defined. Therefore, the system alternatively offered the *information* about all persons (*D_02_06*). Again

the user *requested* the system to name a person with the given name in *D_02_07*, the system found a person with the name, and *requested* the user to confirm, since it is the only person with the last name "Brückner". The system utterance *U_04_08* did not match the user's request in *D_02_07*, showing an unnatural dialogue segment.

D_02_01 U I'd like to go to Brückner.
D_02_02 S Would you like to drive to the doctor Mr. Wolfgang Brückner?
D_02_03 U I need more information about him.
D_02_04 S Mr. Wolfgang Brückner is a doctor in the accident surgery. You find him in room 1711.
D_02_05 U Is there an other person with the name Brückner?
D_02_06 S I found all together 32 persons and they are sorted according to their last names.
D_02_07 U Name a person with the name Brückner.
D_02_08 S Would you like to drive to the doctor Mr. Wolfgang Brückner?
D_02_09 U If there is no other person with the name Brückner, then yes.

5 Dialogue Model Evaluation Concerning Task Success

One aim of our experiment was to investigate the task success of the unified dialogue model in a practical dialogue system. Specifically, in this section we are going to evaluate how the system's deterministic behavior (see Fig. 3(c)) influences the task success. Success at the task of a dialogue in our context is measured by how well the dialogue model supports users to complete dialogue tasks and therefore, we apply the Kappa coefficient [13, 3, 17] approach, similar applications can be found in the literature, such as the evaluation of two train timetable information agents in [17].

First, we define a set of *attribute values* for each task. As shown in Fig. 3(d) the unified dialogue model allows a user to make a dialogue move with an *instruction* like "take me to $\cdots$", a *request* like "tell me about $\cdots$", an *accept* like "yes" or a *reject* like "no" after a system's utterance. Each user's dialogue move may contain some content information, also called *attribute values*, of a person's name, a room number and so on. Tab. 1 summarized the set of all relevant attributes.

Since different tasks contain different data and have different goals, each task has a set of expected dialogue acts and attribute values, such as the *attribute value matrix* (AVM) in Tab. 2 for the task, in which the participants were asked to go to a person with the last name "Brückner" (see the example dialogue *D_02* in Section 4). Each expected dialogue act-attribute pair is associated with an actual value, which reflects the fact that a unified dialogue model contains a state transition based structure at the illocutionary level and an information state management processes. With the attribute value matrix we can develop the confusion matrix for the collected dialogue data of that task (see Tab. 3).

The values in the confusion matrix are obtained by comparing the dialogue moves issued by the participants and the expected attribute values of each task

Table 1. The set of attributes

attribute name	identifier	description	example
first name	FN	first name of a person	Wolfgang
last name	LN	last name of a person	Brückner
gender	G	gender of a person	M
profession	P	profession of a person	Doctor
room number	RNr	number of a room	1711
room type	RT	type of a room	station room
meta room type	MRT	predefined meta type of a room	eating-related
station	F	name of a hospital station	accident surgery

Table 2. An example of value matrices for dialogue acts and attribute values

dialogue act	attribute	actual values
instruct	LN	Brückner
	G	M, F
accept	LN	Brückner
	FN	Wolfgang
	P	Doctor
	G	M

Table 3. An example confusion matrix

data		instruct				accept								other	sum
		LN		G		LN		FN		P		G			
		E	NE	E	NE	E	NE	E	NE	E	NE	E	NE		
instruct	LN	12												4	16
	G			9											9
accept	LN					12									12
	FN							11							11
	P									9					9
	G											11			11

specified by a AVM. A user dialogue move may contain expected or unexpected information with respect to the attribute values defined in the AVM for a dialogue task, so we use "E" and "NE" in confusion matrices to denote such situations. Values in the "other" column record the number of undefined dialogue moves occurred in the dialogue data. Hence, these confusion matrices capture not only expected dialogue situations, but also unexpected and undefined situations.

Given a confusion matrix, the success at reaching dialogue goals is measured with the Kappa coefficient [13, 3, 17]: $\kappa = \frac{P(A)-P(E)}{1-P(E)}$, where $P(A)$ is the proportion of times that the dialogue moves agree with the attribute values and $P(E)$ is the proportion of times that the dialogue moves are expected to be agreed by chance. In our case,

$$P(A) = \frac{\sum_{i=1}^{n} M(i,E)}{T}, \qquad P(E) = \sum_{i=1}^{n} \left(\frac{M(i)}{T}\right)^2$$

where, $M(i,E)$ is the value in an expected column of row i, T is the sum of all user dialogue moves, and $M(i)$ the sum of the user dialogue moves in row i.

Since our goal is to find out how well the dialogue model implemented in the dialogue system supports various types of tasks, instead of individual tasks, we first calculate the Kappa coefficient for each type by the confusion matrix combining all the confusion matrices of the tasks in that type. The first type contains 13 tasks with 149 dialogues, the second type 3 tasks with 35 dialogues, the third type 8 tasks with 88 dialogues. Since the third type contains the second type implicitly, only three tasks were taken in the experiment for the second type. Finally, the three confusion matrices of the three individual task types are combined to a single confusion matrix for computing the total Kappa coefficient. The results are presented in Tab. 4.

Table 4. The task type dependent and independent Kappa coefficients

task type	type I	type II	type III	type I, II, III
Kappa coefficient	$\kappa_1 = 0.99$	$\kappa_2 = 0.85$	$\kappa_3 = 0.82$	$\kappa = 0.94$

6 Discussion of Evaluation Results and Improvements

From the Kappa coefficients calculated in table 4, we can see that the unified dialogue model well supports the three types of dialogue tasks: the Kappa coefficients κ_1, κ_2 and κ_3 are greater than 0.8 and the task type independent Kappa coefficient κ is 0.94.

Especially, the achievement of the dialogue model by the first type of tasks is obvious, since the Kappa coefficient almost equals 1. In all the 13 tasks of the first type the participants were asked to instruct the system to a destination described by several pieces of information about a person, a room or a department (see the dialogue example *D_01* in Section 4), and therefore, the dialogue goal of each task was given explicitly. On the other hand, each dialogue goal in the

third type was described implicitly, such as the goal of the task described by "you are hungry and would like to eat something" was "the patient dining-hall". From the fact $\kappa_1 > \kappa_3$ we can conclude that the unified dialogue model supports dialogues with a clearly defined goal much better than those without. Missing the negotiation of users' intention in the unified dialogue model is the key fact leading to this finding.

The analysis of the confusion matrices issues two drawbacks of the unified dialogue model: losing track of the current user dialogue move type and missing some unexpected dialogue move types. For example, the dialogue *D_03* itself was straightforward. The problem was that the system could not continue the dialogue after the utterance *D_03_03*, since it lost the information to decide whether the utterance is a *request* or an *instruct* move and therefore, the dialogue model cannot determine the next transition to be taken (see Fig. 3(c)). The problem can be solved by adding the *move context* into the information state, which keeps the current user dialogue act, in our case *instruct* or *request*.

D_03_01 U Which profession does Mr. Schulte practice?
D_03_02 S I have found several persons who match Mr. Schulte.
Which person do you mean?
D_03_03 U I mean Mr. Tobias Schulte.

Extending the illocutionary structure of the unified dialogue model is another step to improve the dialogue model. *D_04*, for example, contains the user dialogue move *D_04_02*, where the user told the system that he did not know the room number. This type of dialogue moves is not supported by the illocutionary structure of the dialogue model. Adding a new transition *U.inform* to the start state in the transition dialogue of *Response(U,S)* (see Fig. 3(d)) enables the dialogue model to handle such dialogue moves.

D_04_01 S Would you like to drive to room 1262, ECG 2, in the cardiology?
D_04_02 U I don't know the room number.

The refinements of the dialogue model by adding new information state elements and additional transitions have been applied to update the dialogue system. We believe that they will improve the task success of the unified dialogue model throughout. This has to be proved by a follow-up experiment.

Based on the evaluation results, we conclude that the unified dialogue model well supports users to dialogue with the hospital guidance system, however, they cannot be used to measure the effectiveness of the whole dialogue system, since all the test runs were, with the assistance of a human operator[1], simulated as if the participants were conversing with the system in natural language directly, but removing possible distractions that might have been introduced by speech recognition. Comparing the audio data with the manual input data did not deliver any essential deviation that would affect task successes of any undergone dialogues. Therefore, our focus on evaluation of the unified dialogue model is maintained.

[1] We used only one operator in the whole experiment

Unified dialogue models are constructed at the illocutionary force level, which naturally enables dealing with diversity situations. However, choosing the appropriate set of communicative acts is one important factor affecting the coverage of a unified dialogue model. Care must be taken on the one hand to avoid over-simplification to the point where the structural model collapses down to a two-state initiate-response network with jumps. Although these over-simplified models capture most dialogue situations, they are not useful for dialogue control or formal analysis of dialogue structure. On the other hand, models, as the one discussed in this paper, well reflect natural dialogue structures at the illocutionary level and still possess the context sensitive information state management that relies on domain specific communication. Diversity problems might occur when people dialogue with a system based on a too excessively designed unified dialogue model, but through appropriate design and careful evaluation possible diversities can be detected and the model can then be improved accordingly.

7 Conclusion

In this paper we applied the Kappa coefficient (κ) to evaluate the effectiveness of a unified dialogue model by task success, which combines a generalized dialogic structure at the illocutionary level and an information state based content manager. Specifically, three Kappa coefficients were calculated from the confusion matrices for three types of dialogue tasks using the 272 dialogues collected in an experiment with 12 participants. The results showed that the unified dialogue model well supports those dialogue tasks in general ($\kappa = 0.94$). Especially, tasks with an explicit defined dialogue goal (cf. $\kappa_1 = 0.99$). The experiment results also delivered useful findings for the improvement of the dialogue model. This paper has three major contributions. First, it showed the development of unified dialogue models in general and by an example. Second, we demonstrated how to evaluate unified dialogue models by combining dialogue acts with attribute values. Third, we applied the standard method, the Kappa coefficient, to evaluate a unified dialogue model.

To evaluate the improvement of the unified dialogue model according to the analysis of the experiment results, we are now carrying out a follow-up experiment. The collected dialogue data will also be used for training an automatic speech recognizer, which will then be integrated into the multimodal interactive system for further experimenting. Last but not least, applying reinforcement learning techniques to enhance the existing unified dialogue model centered management system is another research direction we are now concerned with.

Acknowledgments We gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG) through the Collaborative Research Center SFB/TR 8 Spatial Cognition - Subproject I3-SharC. We would like to thank Prof. Dr. Nicole von Steinbüchel, Frank Schafmeister and Nadine Sasse from the Department of Medical Psychology and Medical Sociology at Georg-August-Universität Göttingen for helping us to plan and execute the experiment.

References

1. Failures diference refinement. FDR2 manual. Technical report, Formal System (Europa) Ltd., 2001.
2. J. Allen. *Natural Language Processing*. Benjamin Cumminings Publishing Company Inc., 1995.
3. J. C. Carletta. Assessing the reliability of subjective codings. *Computational Linguistics*, 22(2):249–254, 1996.
4. J. Ginzburg. Dynamics and the semantics of dialogue. *Language, Logic and Computation*, 1, 1996.
5. B. J. Grosz and C. L. Sidner. Attention, Intentions and the Structure of Discourse. *Computational Linguistics*, 12(3):175–204, 1986.
6. C. A. R. Hoare. *Communicating Sequential Processes*. Prentice-Hall, 1985.
7. S. Larsson. *Issue-Based Dialogue Management*. PhD thesis, Department of Linguistics, Göteborg University, 2002.
8. I. Lewin. A formal model of conversational game theory. In *The Fourth Workshop on the Semantics & Pragmantics of Dialogue*, 2000.
9. D. K. Lewis. Scorekeeping in a language game. *Journal of Philosophical Logic*, 8, 1979.
10. A. W. Roscoe. *The Theory and Practice of Concurrency*. Prentice-Hall, 1998.
11. H. Shi, R. Ross, and J. Bateman. Formalising Control in Robust Spoken Dialogue Systems. In B. K. Aichernig and B. Beckert, editors, *Proceedings of Software Engineering and Formal Methods 2005*, IEEE, pages 332–341. IEEE Computer Society, 2005.
12. H. Shi, R. J. Ross, T. Tenbrink, and J. Bateman. Modelling illocutionary structure: Combining empirical studies with formal model analysis. In A. Gelbukh, editor, *Proceedings of the 11th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2010)*, volume 6008 of *Lecture Notes in Computer Science*, pages 340–353, 2010.
13. S. Siegel and N. J. Castellan. *Nonparametric Statistics for the Behavioral Sciences*. McGraw Hill, 1988.
14. S. Sitter and A. Stein. Modelling the illocutionary aspects of information-seeking dialogues. *Journal of Information Processing and Management*, 28, 1992.
15. R. Stalnaker. Assertion. *Journal of Syntax and Semantics*, 9, 1979.
16. D. Traum and S. Larsson. The information state approach to dialogue management. In *Current and New Directions in Discourse and Dialogue*. Kluwer, 2003.
17. M. A. Walker, D. J. Litman, C. A. Kamm, and A. Abella. PARADISE: A framework for evaluating spoken dialogue agents. In *Proc. of the Eighth Conference on European Chapter of ACL*, pages 271–280, 1997.
18. W. Xu, B. Xu, T. Huang, and X. Hairong. Bridging the gap between dialogue management and dialogue models. In *Proc. of the Third SIGdial Workshop on Discourse and Dialogue*, 2002.

Towards Effective, Efficient and Elderly-friendly Multimodal Interaction

Cui Jian
SFB/TR 8 Spatial Cognition, University of Bremen, Germany
ken@informatik.uni-bremen.de

Frank Schafmeister
Medical Psychology and Medical Sociology, Georg August University of Göttingen, Germany
frank-schafmeister@med.uni-goettingen.de

Carsten Rachuy
SFB/TR 8 Spatial Cognition University of Bremen, Germany
rachuy@informatik.uni-bremen.de

Nadine Sasse
Medical Psychology and Medical Sociology, Georg August University of Göttingen, Germany
n.sasse@med.uni-goettingen.de

Hui Shi
SFB/TR 8 Spatial Cognition, University of Bremen, Germany
shi@informatik.uni-bremen.de

Holger Schmidt
Neurology, Georg August University of Göttingen, Germany
h.schmidt@med.uni-goettingen.de

Nicole von Steinbüchel-Rheinwall
Medical Psychology and Medical Sociology, Georg August University of Göttingen, Germany
nvsteinbuechel@med.uni-goettingen.de

ABSTRACT
In this paper we present the design and implementation of a multimodal interactive guidance system for the elderly for the use in hospital environments, which combined common design principles of conventional interactive interfaces and ageing specific characteristics. To evaluate the system we have conducted a pilot study with seven elderly persons. The experiment results are overall positive and therefore support our design decisions. On the other hand, they also reveal some context sensitive problems and advise further improvements.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Evaluation/methodology, Graphical user interfaces (GUI), Haptic I/O, Natural language, Screen design (e.g., text, graphics, color), User-centered design.*

General Terms
Design, Experimentation, Human Factors.

Keywords
Human Computer Interaction, User Centered Design, Elderly Interaction, Multimodal Interaction, Guidance System.

PETRA'11, May 25 - 27, 2011, Crete, Greece.

1. INTRODUCTION
Because of demographic development towards more and more elderly people in today's society (cf. [19]), research in age friendly interactive systems is of increasing importance. In order to facilitate the interaction with modern technical systems while considering the common age-associated decline in physical, cognitive and emotional functions, user interfaces in such systems are mostly multimodal [13]. Research in multimodal interaction for the elderly has focused on various input modalities, (e.g., speech recognition cf. [28], [17]), supportive ambient assisted living environments (cf. [1]), and other. In this paper, we will present an interactive hospital guidance system providing a multimodal interface which combines speech, touch and visual channels. The system provides a basic multimodal interface in which a number of design decisions by elderly persons are implemented.

In order to evaluate the system for further improvements and to study its feasibility and acceptance by the elderly, the effectiveness as well as efficiency by supporting elderly persons in various tasks, a series of experiments with elderly participants has

been planned and two of them are completed. In this paper, the pilot study will be reported, which was conducted with seven participants between the age of 65 and 75 and focused on the combination of inputs via touch screen, speech and visual outputs. This study aimed at the evaluation of the interactive guidance system as a whole, including internal information organization, external presentation and visualization, the mode of interaction, as well as the interaction control mechanism. Specifically, we concentrated on the following system aspects: the effectiveness regarding task success, the efficiency of executing tasks and the user satisfaction regarding the system.

First we will present our design guidelines of the multimodal interface for elderly persons, then we will describe the multimodal interactive guidance system, which is implemented based on conventional interface design principles addressing needs of the elderly to facilitate the use of it. In Section IV the pilot study is presented, and the results are discussed in Section V. Finally in Section VI, we conclude and present our plans for future work.

2. DESIGN GUIDELINES OF MULTIMODAL INTERFACE FOR ELDERLY PERSONS

It's well known that, during normal ageing, decline in sensory, perceptual, motor and cognitive abilities occurs. In the literature (cf.[12],[24],[3]) it is indicated that the different declining processes should be considered for human computer interaction design. Therefore, to address and specify the needs of designing multimodal interface for the elderly, we are presenting a list of guidelines capturing these slowly failing abilities.

These well known constraints guide the designing processes of the effective, efficient and elderly-friendly multimodal interaction. The constraints are introduced in a general way below, then followed by a more practical description regarding the implementation within our multimodal interactive system in the next section.

- **Visual perception** worsens for most people while ageing (cf. [10]). Even in the forties, many people notice it is more difficult to focus on objects up close and to see fine details. The size of the visual field is also decreasing and leads to loss of peripheral vision, rich colors and complicating shapes make images hard or even impossible to identify, and rapidly moving objects are becoming less noticeable. To cope with this decline, several interface design relevant guidelines should be taken into account:
  - The layout of the user interface should be devised as simple and clear as possible, implying few (if any) overlapping items.
  - All texts should be displayed large enough, [12] implying simple fonts lying in the 12-14 point range.
  - Instead of many colors and complicating shapes, few colors while building strong contrast between texts, items and background should be used, as well as simple and easily recognizable designs.
  - Unnecessary and irrelevant animation should be avoided, simple and slow animation can be added only in necessary cases.
- **Speech ability** is essential for the emerging speech enabled interface nowadays. Some authors point out that speech ability declines with ageing due to reduced motor control of tongue and lips and elderly persons often need more time to produce words or longer sentence (cf.[2],[20]). Recent corpus analysis of elderly persons interacting with several spoken dialogue systems however, showed that with elderly-centered adaptation the interaction quality can be improved to a sufficient level (cf.[23],[11]). The following aspects should be noticed while constructing the speech recognizer, analyzer and dialogue management components:
  - Acoustic models specialized for elderly persons should be used for speech recognizer.
  - Vocabulary or grammar of speech analyzer should be strengthened with more definite articles, more auxiliaries, more first person pronouns and more lexical items related to social interaction.
  - Dialogue strategies should be able to cope with elderly centered situations such as repeating, helping, social interaction, etc.
- **Hearing ability** declines with age to 75% for the elderly between 75 and 79 year olds (cf. [10],[15]). High pitched sounds are increasingly missed, as well as long and complex sentences become difficult to follow ([26]). Therefore more attention should be paid to the following:
  - Text displays are needed by the elder persons for misheard information.
  - Synthesized texts should be intensively revised with regard to style, vocabulary and sentence structures of the elderly comparable to the elderlies's speech.
  - Low pitched voices should be used for synthesis, e.g., female voices are less preferred than male.
- **Motor abilities** are also very important while interacting with multimodal interface. Using a computer mouse has been a problem for many elderly because requiring good hand-eye cooperation (cf.[30]). They find it difficult to position the cursor if the target is too small or too irregular to locate (cf.[6]), and have problems with control of fine movement (cf.[30]). In addition, because of the reduced motor functions, more errors can occur during fine movements, especially when other cognitive functions are required at the same time. Therefore, the following points might be noted:
  - Direct interaction via e.g. touch-screen should be recommended.
  - All graphical interface items should be accessibly shaped, sized and well spaced out.
  - Simple movement such as clicking is recommended, and complex movements like dragging, drawing certain shapes should be avoided.
  - Text input should be avoided or replaced with other simpler input actions.
  - Undo or return function should be provided to enable elderly to correct errors or relocate themselves.
  - Simultaneous multimodal input such as the combination of speech and movement input should be avoided and replaced with items requiring only simple movements.
- **Attention** and **Concentration** are reduced with age, elderly persons become more easily distracted by details or noise (cf.[16]), they also have great difficulty to maintain divided

attention where attention must be paid to more than one aspect at the same time (cf.[22]). To cope with these constraints the following are suggested:

- For graphical interfaces only relevant and simple images should be devised and used.
- Unified or similar fonts, colors, sizes of displayed texts are recommended.
- Changes on the user interface should be emphasized in an obvious way.

- **Memory** declines at different degrees for different types. Short term memory holds less items while ageing (cf.[4]), more time is needed to process visual information (cf.[14]). Working memory also becomes less efficient while processing information in short term memory (cf.[25]). Semantic information is believed to be more easily loaded into long term memory (cf.[7]). Prospective memory, the ability to remember, is reduced if complex tasks are involved (cf.[21]). To compensate the decline of different functions, the following points are to be noted:
  - Pure image items should be avoided, or be placed near relevant key words.
  - Presented items should not exceed five, which is the average maximum capacity of the short term memory of elderly people.
  - Presented information should be categorized semantically to assist in memorization into long term memory.
  - Context sensitive information is necessary not only tominimize lightening the load of working memory, but also to remind the elderly of the contextual information.
- **Intellectual ability** does not decline to a lesser extent compared to memory functions in ageing (cf.[27]). However, in [12] it is suggested that crystallized intelligence, the intelligence basing on life experience and solid knowledge, does indicate that elderly people perform best in a stable well-known interface environment. To reflect this on interface design, we suggest to assure:
  - Generally unified interface layout, where changes should only happen on data level.
  - Semantically intuitive structure, where users should not get too surprised while traversing or entering the next levels.
  - Consistent interaction style, which can ease the learning and assist the elderly to master using the interface.

3. MULTIMODAL INTERACTIVE GUIDANCE SYSTEM FOR ELDERLY PERSONS

According to the proposed design guidelines, we developed MIGSEP, a multimodal Interactive Guidance System for elderly. The MIGSEP system runs on a portable touch screen tablet PC, which will serve as an interactive channel on an autonomous intelligent electronic wheelchair that is able to automatically transport elderly or handicapped persons to desired locations while providing relevant information. Therefore, the MIGSEP system provides a general platform for both theoretical researches and empirical studies on multimodal interaction relating to actual application scenarios.

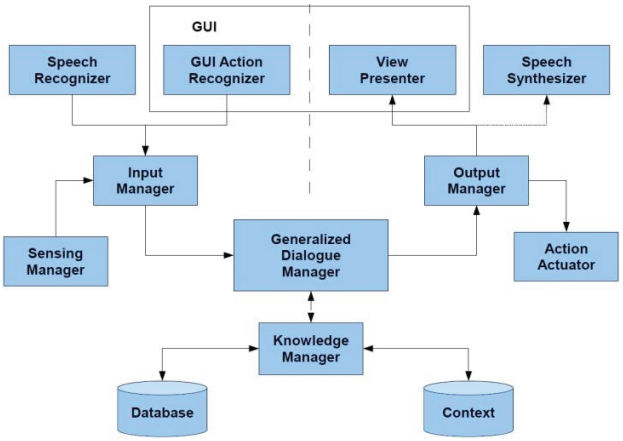


Figure 1. The structure of the MIGSEP system

Figure 1 illustrates the overall structure of MIGSEP system. The Generalized Dialogue Manager is designed by combining generalized dialogue modeling approaches (e.g. [18]) and Information State based dialogue theories (e.g. [29]), and functions as the central processing unit of the entire structure, enables a formally controllable, meanwhile flexible and context-sensitive agent-based dialogue management. The Input Manager collects and interprets all external incoming messages from the GUI Action Recognizer for GUI input events, the Speech Recognizer for natural language understanding and the Sensing Manager for receiving other possible sensor data. The Output Manager on the other hand, handles all outgoing commands and distributes them to the View Presenter for presenting visual feedbacks, the Speech Synthesizer to generate natural language responses and the Action Actuator to perform necessary motor actions. The Knowledge Manager uses the Database to keep the static data of the virtual hospital and the Context to process the dynamic information exchanged with users during the interaction.

Figure 2. User interacting with MIGSEP

Figure 2 shows a user interacting with the MIGSEP system running on a tablet PC. The MIGSEP interface is divided into two areas: a function-area, where the function button start to switch to the start state and the function button back to return to the previous state on the left, the text-presenter to display the system responses in the middle and the function button plan to show the currently planned goals on the right; a choice-area where the information entities are designed as single cards that can be selected and a scrollbar indicates the position of the current shown cards among all while enabling easy jumping to others.

According to the presented guidelines in the previous section, elderly-centered design decisions have been implemented in the MIGSEP system. Specifically, the most essential ones are listed below:

- Visual perception: simple and clear layout as demonstrated in figure 2 has been constructed without overlapping items; 12-14 sized sans-serif font was chosen for all displayed texts. Simple and high contrast colors were used and placed aside, e.g., black blue, black green combinations, etc., as well as regularly shaped items such as rectangles and circles were chosen, enabling comfortably perceived and easily recognized interface elements. Simple and slow animations occur by switching to a new state or an item being selected.
- Hearing ability: a combination of text and acoustic output is provided as system responses. Styles, vocabulary, structures of the sentences have been intensively revised. A low-pitched yet vigorous male voice is chosen for the synthesis.
- Motor functions: regularly shaped, sufficiently sized and well separated interaction elements were designed for easy access. Clicking was decided to be the only action to avoid otherwise frequently occurring errors caused by the declining. Two ways of orientating oneself, start and return, were provided as described above, enabling error corrections and reorientation during interaction.
- Attention or Concentration: fancy and irrelevant images were avoided. The unified font, colors, sizes of interaction elements were used on the entire interface. Simple animation notifying changes were constructed, giving sufficiently clear feedback to the user.
- Memory abilities: all imaged items are combined with relevant keywords. The number of the to be selected items on the main area are restricted to three, considering the maximum capacity of the short term memory, the accessible size as well as the readable amount of information of the interaction items on a table laptop. Logically well-structured items are devised to assist orientation and position during interaction. Context sensitive clues are presented with color changing, e.g. green for person items, yellow for room items, etc.
- Intellectual ability: consistent layout, colors and interaction styles are maintained for easy learning and mastering using the MIGSEP system. Changes on the interface happen only on data level. A semantically intuitive structure closely related to hospital information data was used in order to provide a rather natural environment.

User: <clicks on the blue "Department" card>
Sys: <shows 3 blue cards and a right arrow hinting the other 8 cards>
I've found 11 departments and sorted them alphabetically.
User: <clicks on the blue "cardiology" card>
Sys: <shows 2 cards, green and yellow>
There're two options now: do you want to see all the persons or rooms in the Gastroenterology?
User: <clicks on the green "person" card>
Sys: <shows 3 green cards>
Here are all the persons working in the department of cardiology
User: <clicks on the second green card>
Sys: <enlarges the second green card and shows the yes no buttons>
Do you want to drive to Mr. Dr. Holger Schmidt?
User: <clicks on yes button>
Sys: <resizes the second green card and hide the yes no buttons>
OK, I've saved this goal. Where else do you want to drive?

Figure 3. Example dialogues with MIGSEP

Figure 3 shows a sample dialogue with the MIGSEP system for the assignment ”drive to a person named Holger Schmidt, who is working in the cardiology”, where the user actions and system’s corresponding non-verbal responses are enclosed in angle brackets.

4. Pilot Study

In order to use a multimodal interface, a close cooperation of sensory, perceptual, motor and cognitive functions is required. This is especially difficult for elderly persons who might be suffering a decline of many of those abilities. Therefore, in order to find out how well this merging and management process can be assisted by the MIGSEP system, an evaluation was carried out.

4.1 Participants

Seven elderly persons (five female and two male, with average age of 70 years old, ranging from 65 to 73 with standard deviation 2.91), all German native speakers, took part in the pilot study. All passed the mini-mental state examination (MMSE) (cf.[9]), which is the screening test to measure cognitive mental status, with the average score 29.14 (max 30) (ranging from 27 to 30 with standard deviation 1.07).

4.2 Stimuli and Apparatus

Visual stimuli were displayed on the screen of a portable tablet PC that also generates audio stimuli as part of the feedbacks, shown in figure 2. A direct interaction with the system is then enabled by the touch screen interface of the MIGSEP system.

The same data set, including virtual yet essential information of personnel, rooms and departments in a common hospital, was used throughout the experiment. All tasks were clearly described on paper cards and presented to the participants by an investigator.

In order to collect as many data sets as possible, we used the automatic internal logger of the MIGSEP system, one digital video recorder keeping track of the whole interaction process, a human observer observed the participants and noted all possibly important responses.

A questionnaire concerning user satisfaction degree of the MIGSEP system was also designed, which includes questions of four categories: **screen presentation, system use, system structure** and **task performing**. The questionnaire was completed

by each participant on a four point Likert scale, where one represents the lowest appropriateness and four the highest.

4.3 Procedure

Each participant had to undergo three phases:

- Learning: first a brief introduction was given to the participants, so that they could acquire the basic idea and overview of the experiment, the way how to interact with MIGSEP system and also get an insight in the verbal and graphical feedbacks the system provides. Then they were asked to perform several sample tasks to gain a deeper insight into the system and practical experiences.
- Testing: Each participant had to perform ten tasks, each of which contains incomplete yet sufficient information (e.g. the room number or the first name of a person) about a destination the participant should select, e.g., to drive to the room 1261, or drive to a person named Wolf in the Ear-Nose-Throat department. Each task was finished, if either the goal was selected, or the participant gave up trying.

Evaluation: After the ten tasks were completed, each participant was asked to fill in the evaluation questionnaire.

4.4 Study Questions and Methods

Our first hypothesis concerns the question ”Can the elderly use the system to complete the tasks?”, i.e., the effectiveness of the MIGSEP interface was evaluated. In this study we use the standard effectiveness-measurement method Kappa coefficient (cf. [5] and [31]) to assess the successfulness of the interaction between the elderly and the system.

The second aspect being considered is the efficiency of the interaction: Can the elderly users handle the tasks with the system efficiently? This was answered by analyzing the automatic logged data for every single interaction step with the MIGSEP system.

The next question to be tested with the MIGSEP interface was: Whether the elderly find it comfortable interacting with MIGSEP, i.e., whether our age tailored design assisted the elderly or not. This should be reflected in comparing the results of the questionnaires that the participants were asked to fill after the interaction.

Finally, there is an open question we wish to investigate: ”Is there anything which is wrong, unnecessary or missing?” This should be answered by combining the objective observations during the interacting process and the subjective comments about the pointed questions raised after the interaction.

5. RESULTS, ANALYSIS AND DISCUSSION

Regarding the four questions raised in the previous section, we are going to demonstrate, analyze and discuss about the experiment data in this section.

5.1 Effectiveness of the Interaction

The first question we want to find out through the experiment is, how well the MIGSEP interface assists the elderly persons to perform the tasks, i.e., the effectiveness of the elderly-centered designed interaction. Kappa coefficient is a well accepted method for effectiveness measurement (cf. [5] and [31]). In order to apply this method, we need to define the attribute value matrix (AVM), which contains all information that has to be exchanged between MIGSEP and the subjects to accomplish the given tasks. The attributes used throughout the experiment are listed in table I.

Table I. The Attributes of AVMs for Kappa Coefficient

Attribute name	Identifier	example
First name	FN	Ken
Last name	LN	Fischer
Gender	G	Male
Profession	P	Doctor
Room number	RNr	1711
Room name	RN	Message room 2
Meta room type	MRT	Eating-related
Department	D	Accident surgery

With the listed attributes we can construct the AVMs for all the tasks, e.g. table II shows the AVM for the task: ”to drive to a person named Wolf in the department of Otolaryngology”, where the expected attribute value pairs of this task are presented.

Table II. The Example AVM for the Task “Drive to a person named Wolf in otolaryngology.”

Attribute	Actual values
FN	Diana
LN	Wolf
G	Female
D	Otolaryngology

By comparing the actual data collected in the experiment with the expected attribute value pairs in the AVMs, we can construct the confusion matrices for different tasks, e.g., table III for ”drive to a person named Wolf in Otolaryngology”, where ”M” and ”N” are used to denote whether the actual data match with the expected attribute values in the AVMs or not, e.g. one participant picked the wrong department as shown in table III.

Table III. The Confusion Matrix for the Task “Drive to a person named Wolf in otolaryngology”

	FN		LN		G		D		sum
data	M	N	M	N	M	N	M	N	
FN	6								6
LN			6						6
G					6				6
D							6	1	7

Given one confusion matrix, the Kappa coefficient can be calculated with $\kappa = \frac{P(A)-P(E)}{1-P(E)}$, (cf. [5] and [31]). In our experiment, $P(A) = \frac{\sum_{i=1}^{n} M(i,\ M)}{T}$ is the proportion of times that the actual data agree with the attribute values, and $P(E) = \sum_{i=1}^{n} (\frac{M(i)}{T})^2$ is the proportion of times that the actual data are expected to be agreed by chance, where M(i, M) is the value of the matched cell of row i, M(i) the sum of the cells of row i, and T the sum of all cells. Therefore, we summarized the results of all the tasks and construct one overall confusion matrix, and got that,

P(A) = 0.971 and kappa coefficient $\kappa = 0.966$, suggesting a highly successful degree of the interaction between MIGSEP system and the participants.

5.2 Efficiency of the Interaction

Table IV summarizes the quantitative results of the 7 participants with respect to user turns, system turns and elapsed time. Averagely 8.06 user turns and 5.94 system turns per task show very good performance efficiency for the elder persons, because the average minimal turn numbers, which can be inferred by analyzing the shortest solution of each task, are 7.1 user turns and 5.3 system turns, suggesting almost every participant was able to pick the fastest way to perform tasks.

Table IV. Quantitative Results Concerning Efficiency

	User turns	Sys turns	Elapsed time (s)
P1	8.0	6.3	105.2
P2	8.8	6.4	54.8
P3	7.3	5.5	45.7
P4	8.6	6.1	67.1
P5	8.6	6.0	71.7
P6	7.6	5.5	38.2
P7	7.5	5.8	51.8
Avg.	8.06	5.94	62.07
Stdv.	0.57	0.33	20.61

Although the average elapsed time also shows satisfying results, with averagely 62.07 second per task for minimal 12.4 interaction paces (7.1 user turns + 5.3 system turns), the standard deviation 20.61 is considered too high. The first reason that can be observed directly from Table IV is the elapsed time of the participant P1, who in fact either gave wrong solutions or failed to complete 3 out of 10 tasks.

After a thorough analysis of P1's data, the major problem was the successfulness while performing tasks, but the function "Return". "Return" is proved to be very helpful on conventional user interfaces, e.g., in nowadays web browsers, enabling jumping back to the previous state. However, this function was not frequently used by the seven participants, P1, e.g., only used the function twice, then spent much more time to orientate himself and got lost in the end. Same problems happened to the second lowest, 71.7 second for using return once and being completely lost while performing one task. Even though there seems to be one exception, the participant P2, who always used the return function throughout the experiment and yet spent only averagely 54.8 second performing each task. But after a revisit of P2's data, it was noted that P2 used return only for going back to the start state of the system, which can be done alternatively with one click on the start button, causing a waste of user/system turns (P2 had the most user and system turns). Therefore, we can conclude that, presumably due to the declines in memorization ability, "Return" function is a harming factor for building efficient interaction with elderly persons.

5.3 User Satisfaction

Our next concern lies in the assessment of subjective user satisfaction. Table V shows the summarized result drawn from the evaluation questionnaire. On the whole, the subjective user satisfaction degree is very good with the score of 3.39 out of 4. Specifically, the presentation of the interface is intuitive and easy to understand, this can be observed that users rated the screen presentation with the score of 3.32; the use of the system has been feasible, noted from the score of 3.39; the structure design of the system is also considered as reasonable by the users, having the score of 3.43; and finally the users found it well-assisted performing the task, by giving the score of 3.45.

Table V. Results of the Assessment of Subjective User Satisfaction

Aspect	Score	Stdv.
Screen Presentation	3.32	0.33
System Use	3.39	0.38
System Structure	3.43	0.29
Task Performing	3.45	0.28
Overall	3.39	0.05

On the other hand, the standard deviations of the screen presentation and system use are a bit higher, with 0.33 and 0.38 respectively, implying greater differences coming from some users, this is presumably because of the grading relating to the following questions:

- [Category: screen presentation] Do the colors of different elements on screen ease the understanding of the presentation?

  The standard deviation of the rating is 0.69, showing that some users had problems noticing the color differences during the interaction. This is probably caused by a major decline in the visual ability. But meanwhile the other questions concerning the screen presentation have been given good ratings, suggesting that the participants may have already been assisted by the colors or combination with others unconsciously.

- [Category: system use] Is it easy to use the system? do you only need a short time for learning to use the system?

  The standard deviations are 0.76 and 0.69, respectively, indicating that some users found it unpleasant to use or learn to use the system. This may relate to declines of corresponding abilities and the management of all necessary ones, but it should also be noted that many participants only have rather little experience using computer devices and might therefore have problems to learn, to use and to express their opinions concerning the system.

5.4 Are we doing the right things?

In order to answer this question, the notes of objective observations during the interaction and the subjective comments about some pointed question after the interaction are analyzed and summarized as follows.

All participants found it convenient and easy to understand and use the MIGSEP system. Specifically, the comforting and high contrast colors, the appropriately sized and designed texts and elements, the explicitly reduced yet sufficient information, etc.,

are highly appraised. Most participants also expressed that even without participating in the tutorial before the testing phase would not have any big problem in using the system, only a little longer time could be needed to get familiar with the system. One even emphasized that she could not use the ticket machine on bus to buy a ticket, but was able to interact with the system to perform all the tasks.

On the other hand, relevant suggestive opinions were also given. The first suggestion indicates that, more context information can be hinted at assisting the elderly persons to orientate themselves after system state is switched to a new one, e.g., if the department is selected, the name of the department can be shown in the next state telling the participants where they are. The next suggestion implies that, the logically well-organized structure is less preferred than more intuitive and direct ones for elderly persons, this has been mentioned by some participants as they found some common or frequently used facilities should be made more easily accessible, e.g., the toilets, regardless the fact that ”toilets” are too specific to be put at the first logical level under location. Furthermore, it could be observed during the experiment that, elderly persons tended to rely on simpler actions or feedbacks, suggesting that e.g. pressing actions are easier to performed than the conventional released actions on interacting elements, or locating with as well as receiving feedbacks from a moving scrollbar do not provide extra helps as they usually do, and even produce more problems because of the elderly persons' declines in visual and motor abilities.

6. CONCLUSION AND FUTURE WORK

In this paper we presented our work on multimodal interaction for elderly persons from two perspectives: the design and implementation of a multimodal interactive guidance system for elderly persons according to a number of guidelines that combine the basic design principles of conventional interactive interfaces and the most common ageing centered characteristics; and the evaluation of the system with seven elderly persons. The evaluation findings showed that the system is effective, efficient and has a high user satisfaction. Therefore we gained further evidence for our proposed guidelines on system design and implementation. In addition, the experimental results also enables us to improve several functions of the system.

The reported work served as a first step of a series of planned development processes towards building effective, efficient, adaptive and robust multimodal framework(s) for interaction between human operators and service robots. The second experiment concerning speech modality has also been completed and the results are being analyzed, further experiments are being conducted as well. Moreover, reinforcement learning techniques will be applied to gain more flexible and powerful interaction. Our future researches will focus on the combination of speech, touch and visual modalities, discourse modeling and management and reinforcement learning in advanced multimodal interactive systems for elderly persons.

7. ACKNOWLEDGMENTS

We gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG) through the Collaborative Research Center SFB/TR8 – Project I3-[SharC] “Shared Control via Interaction” and the department of Medical Psychology and Medical Sociology of Georg August University of Göttingen

8. REFERENCES

[1] Abascal, J. and Castro, I.F.De and Lafuente, A. and Cia, J.M., Adaptive interfaces for supportive ambient intelligence environments. *In Proceedings of the 11th International Conference on Computers Help People with Special Needs*, 2008.

[2] Balota, D.A. and Duchek, J.M., Age-related differences in lexical access, spreading activation and simple pronunciation. *Psychology and Aging 3*, pages 84–93, 1988.

[3] Birdi, K. and Pennington, J. and Zapf, D., Aging and errors in computer based work: an observational study. *Journal of Occupational and Organizational Psychology*, pages 35–74, 1997.

[4] Botwinick, J. and Storandt, M.,(eds.). Memory Related Functions and Age. Charles C Thomas Pub Ltd, 1974.

[5] Carletta, J.C., Assessing the reliability of subjective codings. *Computational Linguistics*, 22(2):249–254, 1996.

[6] Charness, N. and Bosman, E., Human factors and design. In J. E. Birren and K. W. Schaie, (eds.), *Handbook of the Psychology of Aging*, volume 3, pages 446–463. Academic Press, 1990.

[7] Craik, F. and Jennings, J., Human memory. In F. Craik and T. A. Salthouse, (eds.), *The Handbook of Aging and Cognition*, pages 51–110. Erlbaum, 1992.

[8] Dixon, R.A. and Kurzman, D. and Friesen, I., Handwriting performance in younger adults: Age, familiarity and practice effects. *Psychology and Aging 8*, pages 360–370, 1993.

[9] Folstein, M. and Folstein, S. and McHugh, P., ”mini-mental state”. a practical method for grading the cognitive state of patients for clinician. *Journal of psychiatric research*, 12(3):189–198, November 1975.

[10] Fozard, J. L., Vision and hearing in aging. In J. Birren, R. Sloane, and G. D. Cohen, (eds.), *Handbook of Mental Health and Aging*, volume 3, pages 150–170. Academic Press, 1990.

[11] Georgila, K. and Wolters, M. and Karaiskos, V. and Kronenthal, M. and Logie, R. and Mayo, N. and Moore, J. and Watson, M., A fully annotated corpus for studying the effect of cognitive ageing on users' interactions with spoken dialogue systems. *In Proceedings of the 6th International Conference on Language Resources and Evaluation*, 2008.

[12] Hawthorn, D., Possible implications of ageing for interface designers. *Interacting with Computers*, pages 507–528, 2000.

[13] Holzinger, A. and Mukasa, K.S. and Nischelwitzer, A.K., Introduction to the special thematic session: Human-computer interaction and usability for elderly. *In Proceedings of the 11th International Conference on Computers Help People with Special Needs*, 2008.

[14] Hoyer, W. J. and Rybash, J. M., Age and visual field differences in computing visual spatial relations. *Psychology and Aging 7*, pages 339–342, 1992.

[15] Kline, D. W. and Scialfa, C. T., Sensory and perceptual functioning: basic research and human factors implications. In A. D. Fisk and W. A. Rogers, (eds.), *Handbook of Human Factors and the Older Adult.* Academic Press, 1996.

[16] Kotary, L. and Hoyer, W. J., Age and the ability to inhibit distractor information in visual selective attention. *Experimental Aging Research*, 1995.

[17] Krajewski, J.and Wieland, R. and Batliner, A., An acoustic framework for detecting fatique in speech based human computer interaction. *In Proceedings of the 11th International Conference on Computers Help People with Special Needs*, 2008.

[18] Lewin, I. and Lane, M., A formal model of conversational game theory. *In Proc. Gotalog-00, 4th Workshop on the Semantics and Pragmatics of Dialogue*, Gothenburg, 2000.

[19] Lutz, W. and Sanderson, W. and Scherbov, S., The coming acceleration of global population ageing. *Nature*, pages 716–719, 2008.

[20] Mackay, D. and Abrams, L., Language, memory and aging. In J. E. Birren and K. W. Schaie, editors, *Handbook of the Psychology of Aging*, volume 4, pages 251–265. Academic Press, 1996.

[21] McDaniel, M.A. and Einstein, G.O., The importance of cue familiarity and cue distinctiveness in prospective memory. *Memory 1*, pages 23–41, 1993.

[22] McDowd, J.M. and Craik, F., Effects of aging and task difficulty on divided attention performance. *Journal of Experimental Psychology: Human Perception and Performance 14*, pages 267–280, 1988.

[23] Moeller, S. and Goedde, F. and Wolters M., Corpus analysis of spoken smart-home interactions with older users. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odjik, S. Piperidis, and D. Tapias, (eds.), *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, may 2008. European Language Resources Association (ELRA).

[24] Morris, J.M., User interface design for older adults. *Interacting with Computers*, pages 373–393, 1994.

[25] Salthouse, T.A. The aging of working memory. *Neuropsychology 8*, pages 535–543, 1994.

[26] Schieber, F., Aging and the senses. In J. E. Birren, R. B. Sloane, and G. D. Cohen, (eds.), Handbook of Mental Health and Aging, volume 2. Academic Press, 1992.

[27] Shaie, K. W., Intellectual development in adulthood. In J. E. Birren and K. W. Shaie, (eds.), *Handbook of the psychology of aging*, volume 4. Academic Press, 1996.

[28] Takahashi, S. and Morimoto, T. and Maeda, S. and Tsuruta, N., Dialogue experiment for elderly people in home health care system. *In Proceedings of the 6th International Conference on Text, Speech and Speech*, 2003.

[29] Traum, D. and Larsson, S., The information state approach to dialogue management. In J. van Kuppevelt and R. Smith, (eds.), *Current and New Directions in Discourse and Dialogue*, pages 325–354. Kluwer, 2003.

[30] Walkder, N. and Philbin, D.A. and Fisk, A.D., Age-related differences in movement control: adjusting submovement structure to optimize performance. Journal of Gerontology: *Psychological Sciences 52B*, pages 40–52, 1997.

[31] Walker, M.A. and Litman, D.J. and Kamm, C.A. and Kamm, A. A. and Abella, A., Paradise: A framework for evaluating spoken dialogue agents. pages 271–280, 1997

EVALUATING A SPOKEN LANGUAGE INTERFACE OF A MULTIMODAL INTERACTIVE GUIDANCE SYSTEM FOR ELDERLY PERSONS

Cui Jian[1], Frank Schafmeister[2], Carsten Rachuy[1], Nadine Sasse[2], Hui Shi[1,3], Holger Schmidt[4] and Nicole von Steinbüchel[2]

[1]*SFB/TR8 Spatial Cognition, University of Bremen, Enrique-Schmidt-Straße 5, Bremen, Germany*
[2]*Medical Psychology and Medical Sociology, University Medical Center Göttingen, Waldweg 37, Göttingen, Germany*
[3]*German Research Centre for Artificial Intelligence, University of Bremen, Enrique-Schmidt-Straße 5, Bremen, Germany*
[4] *Neurology, University Medical Center Göttingen, Robert-Koch-Str. 40, Göttingen, Germany*
ken@informatik.uni-bremen.de, frank-schafmeister@med.uni-goettingen.de, rachuy@informatik.uni-bremen.de, n.sasse@med.uni-goettingen.de, hui.shi@dfki.de, {h.schmidt, nvsteinbuechel}@med.uni-goettingen.de

Keywords: ICT and ageing, elderly-friendly interaction, user centered design, human-computer interaction, spoken dialogue systems, formal methods, multimodal interaction

Abstract: This paper presents a multimodal interactive guidance system for elderly persons for the use in navigating in hospital environments. We used a unified modelling method combining the conventional recursive transition network based approach and agent-based dialogue theory to support the development of the central dialogue management component. Then we studied and specified a list of guidelines addressing the needs of designing and implementing multimodal interface for elderly persons. As an important step towards developing an effective, efficient and elderly-friendly multimodal interaction, the spoken language interface of the current system was evaluated by an elaborated experiment with sixteen elderly persons. The results of the experimental study are overall positive and provide evidence for our proposed guidelines, approaches and frameworks on interactive system development while advising further improvements.

1 INTRODUCTION

Multimodal interfaces are becoming more and more common since the inspirational introduction by (Bolt, 1980). They are considered as a promising possibility to improve the quality of communication between users and systems and have significant impact on effectiveness and efficiency of interaction (cf. e.g. (Jaimes and Sebe, 2007)), they also enhance users' satisfaction and provide a more natural and intuitive way of interaction (cf. e.g. (Oviatt, 1999)).

Meanwhile, the demographic development towards more elderly keeps motivating the research of elderly-friendly interactive systems; there is a special focus on the multimodal communication channels, which can enhance interaction by taking age-related decline into special accounts (Holzinger, Mukasa and Nischelwitzer, 2008).

In this paper, we will present an interactive guidance system for elderly persons. It uses a unified dialogue modelling approach combining the classic agent based dialogue theories and a formal language supporting generalized recursive transition network based method to achieve a flexible and context-sensitive, yet formally tractable and controllable interaction. Furthermore, it is developed according to a number of elaborated guidelines regarding basic design principles of conventional interactive systems and most common elderly-centered characteristics. To evaluate this system with respect to its feasibility and acceptance by elderly, an experimental study was conducted, which was focused on the natural spoken language input interface of the system. However, the study also aimed at evaluation of the multimodal interactive guidance system as a whole, while regarding the essential criteria of the following aspects: the effectiveness of task success, the efficiency of executing tasks and the user satisfaction with the system.

The remainder of the paper is organized as follows: section 2 introduces the formal unified dialogue modelling approach which combines the

classic agent based approach and the recursive transition network based theory for building the discourse management of the multimodal interaction; section 3 presents a set of specific guidelines for designing multimodal interactive system for elderly persons; section 4 then describes the multimodal interactive guidance system, which is developed based on the unified dialogue modelling approach and the proposed set of design guidelines; in section 5 the experiment is described, and the results are analysed and discussed in section 6. Finally, in section 7 we will conclude and give an outline of future work.

2 A FORMAL UNIFIED DIALOGUE MODELLING APPROACH

As a typical recursive network based approach, generalized dialogue models were developed by constructing dialogue structures at the illocutionary level (Sitter and Stein, 1992). However, it is criticized for its inflexibility of dealing with dynamic information exchange. On the other hand, information state update based theories were deemed the most successful foundation of agent based dialogue approaches (Traum and Larsson, 2003), which provides a powerful mechanism to handle dynamic information and gains a context sensitive dialogue management. Nevertheless, such models are usually very difficult to manage and extend (Ross, Bateman and Shi, 2005).

Thus, a unified dialogue modelling approach was developed. It combines the generalized dialogue models with information state updated based theories. This approach is supported by a formal development toolkit, which is used to implement an effective, flexible, yet formally controllable dialogue management for multimodal interaction.

2.1 A Unified Dialogue Modelling Approach

Generalized dialogue models can be constructed with the recursive transition networks (RTN). They abstract dialogue models by describing illocutionary acts without reference to direct surface indicators (Alston, 2000). Figure 1 shows a simple generalized dialogue model as a recursive transition network diagram. It is initiated with an assertion from a person A, and responded by B with three possible actions: accept, agree or reject.

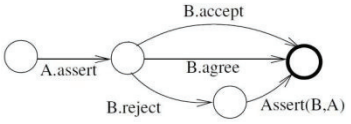


Figure 1: a generalized dialogue model as a simple RTN.

The generalized dialogue model above is a none-deterministic model, to build a feasible interaction model, deterministic behaviour should be assured for the interaction flow. Thus, conditional transitions are introduced to modify the above dialogue model (cf. figure 2). Let *checkAssert* be a method to check whether an assertion holds with B's knowledge and *a* an assertion given by A, if the assertion holds, B can agree with it; otherwise, B rejects it and initiates further discussion; if the assertion is not known by B, then B accepts it. Such conditional transitions can only be activated if the relevant condition is fulfilled. We call it the conditional RTN.

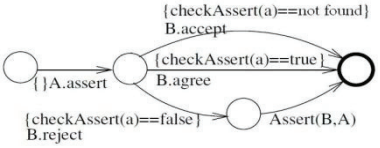


Figure 2: a generalized dialogue model as a simple deterministic RTN with conditional transitions.

Although the conditional RTN based generalized dialogue model defines a deterministic illocutionary structure, it does not provide the mechanism to integrate discourse information. Thus, information state based theory was integrated into our unified dialogue model by eliminating some typical elements, e.g. AGENDA for planning the next dialogue moves, because such information is already captured by the generalized dialogue model; furthermore it complements illocutionary structure with update rules, which is associated with the information state of current context, and can update the information state respectively if necessary. As a result, a unified dialogue model is constructed as shown in figure 3. Four update rules are added, so that the information state regarding context can always be considered and updated; e.g. the update rule ACCEPT is used to add a new assertion *a* into B's belief and refer it as known from then on.

Finally, we define a unified dialogue model as a deterministic recursive transition network built at the illocutionary level of interaction processes; its transitions can only be triggered by fulfilled conditions concerning the information state, and with the consequences of possible information state update according to a set of update rules.

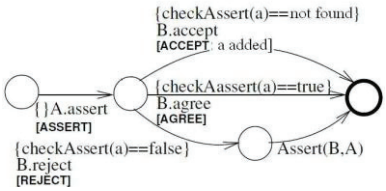


Figure 3: unified dialogue model as a simple deterministic RTN with conditional transitions and update rules.

2.2 A Formal Language Based Development Toolkit for Dialogue Modelling

Deterministic recursive transition networks can be illustrated as a typical finite state transition diagram (cf. figure 3), which provides the possibility of specifying the described illocutionary structure with mathematically well-founded formal methods, e.g., with *Communicating Sequential Processes* (CSP) in the formal methods community of computer science.

CSP can not only be used to specify finite state automata structured patterns with abstract, yet highly readable and easily maintainable logic formalization (cf. (Roscoe, 1997)), but it is also supported by well-established model checkers to verify the concurrent aspects and increasing the tractability (Hall, 2002). Thus, CSP is used to specify and verify the unified dialogue models (cf. the example in figure 4).

```
UDM = A.assert -> B.checkAssert ->
        (  B.accept -> UDM
        [] B.agree -> UDM
        [] B.reject -> AssertBA)
```

Figure 4: a sample CSP specification of the illocutionary structure of the unified dialogue model in figure 3.

In order to support the development of unified dialogue models within practical interactive systems, we provided FormDia, the Formal Dialogue Development Toolkit (cf. figure 5).

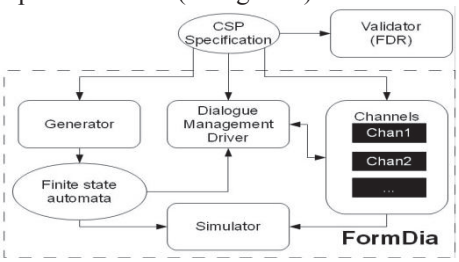


Figure 5: the Structure of the FormDia Toolkit (cf. (Shi and Bateman, 2005)).

To develop the unified dialogue model based management, FormDia toolkit can be used according to the following essential steps:

- **Validation**: the CSP specified structure of a unified dialogue model can be validated by using Failures-Divergence Refinement tool, abbrv. FDR (Broadfoot and Roscoe, 2000), which is a model checking tool for validating and verifying concurrency of state automata.
- **Generation**: according to the given CSP specification, finite state automata can then be generated by the FormDia Generator.
- **Channels Definition**: channels between the dialogue management and application/domain specific components can be defined. These channels are at first black boxes, which will later be filled with deterministic behaviour of concrete domain components.
- **Simulation**: with the generated finite state automata and the communication channels, dialogues scenarios are simulated via a graphical interface, which visualizes dialogue states as a directed graph and provides a set of utilities to trigger events and the dialogue state update for testing and verification.
- **Integration**: after the dialogue model is validated, tested and verified, it can be directly integrated into a practical interactive dialogue system via a dialogue management driver.

The FormDia toolkit shows a promising way for developing formally tractable and extensible interaction. It enables an intuitive design of dialogue models with formal language, automatic validation of related functional properties, and it also provides an easy simulation, verification for the specified dialogue models, and the straightforward integration within a practical interactive system. In addition, with the unified dialogue model, FormDia toolkit can even be used in multimodal interactive system.

3 DESIGN GUIDELINES OF MULTIMODAL INTERACTIVE SYSTEMS FOR ELDERLY PERSONS

Elderly persons often suffer from decline of sensory, perceptual, motor and cognitive abilities due to age-related degenerative processes. (Birdi, Pennington and Zapf, 1997) and (Morris, 1994) indicated that this decline should be considered while designing interactive systems for the elderly. Therefore, we defined a set of design guidelines for multimodal interaction with respect to the decline of the most common abilities. They are implemented and integrated into our multimodal interactive guidance

system and tested by a pilot study. The results are described in (Jian, et al., 2011) and the improved guidelines are now presented as follows, regarding the decline of the seven most common abilities.

3.1 Visual Perception

Visual perception declines for most people with age (Fozard, 1990). Even in the early forties, many people find it more difficult to focus on objects up close and to see fine details. The size of the visual field is decreasing and leads to loss of peripheral vision. Rich colours and complex shapes make images hard or even impossible to identify. Rapidly moving objects are either causing too much distraction, or becoming less noticeable. To cope with these impairments, the following guidelines should be taken into account:

- Layouts of the user interface should be devised as simple and clear as possible, with few (if any) or no overlapping items.
- All texts should be large enough, suggesting simple fonts in the 12-14 point range.
- Strong contrast should be used with as few colors as possible; this also applies to simple and easily recognizable shape designs.
- Unnecessary and irrelevant visual effects and animation should be avoided.

3.2 Speech ability

Elderly persons need more time to produce complex words or longer sentences, probably due to reduced motor control of tongue and lips (Mackay and Abrams, 1996). Furthermore, speech-related elderly-centered adaptation is necessary to improve the interaction quality to a sufficient level (Moeller, Goedde and Wolters, 2008). Based on these, the following aspects should be taken into account:

- Acoustic models specialized for the elderly should be used for speech recognizer.
- Vocabulary should be built with more definite articles, auxiliaries, first person pronouns and lexical items related to social interaction.
- Dialogue strategies should be able to cope with elderly specific needs such as repeating, helping and social interaction, etc.

3.3 Hearing ability

Hearing ability declines to 75% with increasing age 75 and 79 year olds, (Kline and Scialfa, 1996). High pitched sounds are increasingly not percieved, as well as long and complex sentences becoming difficult to follow (Schieber, 1992). Therefore special attention should be paid to the following:

- Text displays can help when information is mis- or not heard.
- Synthesized texts should be intensively revised regarding style, vocabulary, length and sentence structures suitable for elderly.
- Low pitched voices are more acceptable for speech synthesis, e.g., female voices are less preferred than male ones.

3.4 Motor abilities

Using a computer mouse has been problematic for many elderly persons as good hand-eye coordination is required (Walkder, Philbin and Fisk, 1997). It is difficult for them to position the cursor if the target is too small or too irregular to locate, and they have problems with control of fine movements (Charness and Bosman, 1990), especially when other cognitive functions are required at the same time. Thus, the following procedures are suggested:

- Direct interaction is recommended.
- All GUI items should be accessibly shaped, sized and well spaced from each other.
- Simple movements are recommended, such as clicking instead of dragging or drawing.
- Text input should be avoided or replaced with other simpler input actions.
- An undo function is needed to correct errors.
- Simultaneous multimodal input such as the combination of speech and other input should be avoided or replaced.

3.5 Attention and Concentration

Elderly individuals become more easily distracted by details or noise (Kotary and Hoyer, 1995). They display great difficulty maintaining divided attention, e.g. where attention must be paid to more than one aspect at the same time (McDowd and Craik, 1988). To cope with these constraints the following points are suggested:

- Only relevant images should be used.
- Items should not be displayed simultaneously.
- Unified or similar fonts, colors and sizes of displayed texts are recommended.
- Changes on the user interface should be emphasized in an obvious way.

3.6 Memory

Different memory functions decline at different degrees during ageing. Short term memory holds fewer items while ageing and more time is needed to process information (Hoyer and Rybash, 1992). Working memory also becomes less efficient (Salthouse, 1994). Semantic information is believed to be preserved in long term memory (Craik and Jennings, 1992). To compensate the decline of the different memory functions, the following points are suggested:

- Pure image items should be avoided or placed near relevant key words.
- Presented items should not exceed five, the average maximum capacity of short term memory of elderly people.
- Information should be categorized to assist storage into long term memory.
- Context sensitive information is necessary to facilitate working memory activities.

3.7 Intellectual ability

Fluid intelligence does decline with ageing (Shaie, 1996), however, crystallized intelligence does not (Hawthorn, 2000); it can assist elderly people to perform better in a stable well-known interface environment. To reflect this on interface design, we suggest assuring the following points:

- Unified interface layout, where changes should only happen on data level.
- Semantically intuitive structure, where users should not be too surprised while traversing the interaction levels.
- Consistent interaction style, easing learning and assist elderly to master interface use.

4 MULTIMODAL INTERACTIVE GUIDANCE SYSTEM FOR ELDERLY PERSONS

The Multimodal Interactive Guidance System for Elderly Persons (MIGSEP) was developed for elderly or handicapped persons to navigate through public spaces. MIGSEP runs on a portable touch screen tablet PC. It serves as the interactive media designed for an autonomous intelligent electronic wheelchair that can automatically carry its users to desired locations within complex environments.

4.1 System Architecture

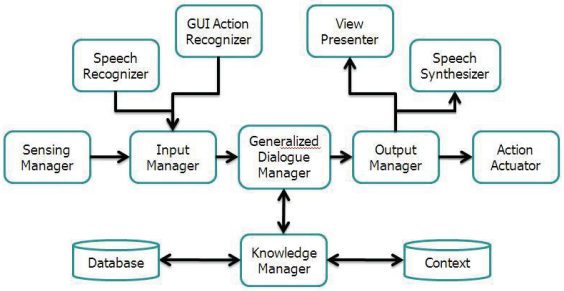


Figure 6: The architecture of MIGSEP.

The architecture of MIGSEP is illustrated in figure 6. A *Generalized Dialogue Manager* is developed using the unified dialogue modelling approach. It functions as the central processing unit and enables a formally controllable and extensible, meanwhile context-sensitive multimodal interaction. An *Input Manager* receives and interprets all incoming messages from *GUI Action Recognizer* for GUI inputs, *Speech Recognizer* for natural language understanding and *Sensing Manager* for other sensor data. An *Output Manager* on the other hand, handles all outgoing commands and distributes them to *View Presenter* for visual feedbacks, *Speech Synthesizer* to generate natural language responses and *Action Actuator* to perform necessary motor actions. *Knowledge Manager* uses *Database* to keep the static data of certain environments and *Context* to process the dynamic information exchanged with users during the interaction.

Although the essential components of MIGSEP are closely connected with each other via predefined XML-based communication mechanism, each of them is treated as an open black box and can be implemented or extended for specific use, without affecting other MIGSEP components. It provides a general platform for both theoretical researches and empirical studies on multimodal interaction.

4.2 The Unified Dialogue Model in MIGSEP

The current unified dialogue model (UDM) consists of four extended state transition diagrams.

The interaction is initiated with the diagram *Dialogue(S, U)* (cf. figure 8), by the initialization of the system's start state and a greeting-like request.

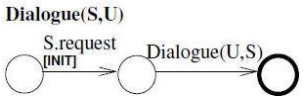


Figure 8: the initiate diagram.

The dialogue continues with user's instruction to a certain location, request for a certain information or restart action, leading to the system's further response or dialogue restart, respectively, as well as updating the information state with the attached update rules (cf. *Dialogue(U, S)* in figure 9).

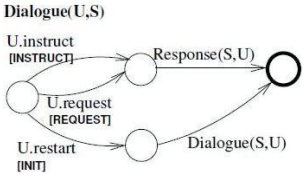


Figure 9: the transition diagram triggered by the user.

After receiving user's input, the system tries to generate an appropriate response according to its current knowledge base and information state (cf. *Response(S, U)* in figure 10). This can be informing the user with requested data, rejecting an unacceptable request with or without certain reasons, providing choices for multiple options, or asking for further confirmation of taking a critical action, each of which triggers transitions to different diagrams.

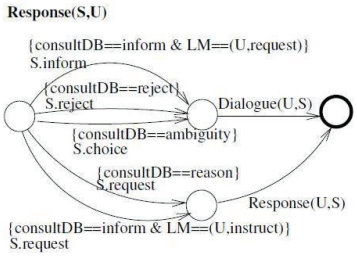


Figure 10: the system's response.

Finally, the user can accept or reject the system's response, or even ignore it by simply providing new instructions or requests, triggering further state transitions as well as information state updates (cf. *Response(U, S)* in figure 11).

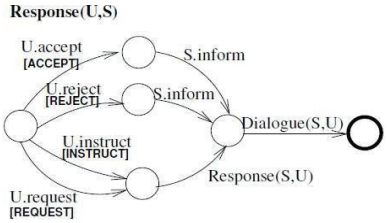


Figure 11: the user's response.

Using the FormDia toolkit, the UDM was developed as CSP specifications, and its functional properties have been validated and verified via FDR, as well as its conceptual interaction process using FormDia simulator. The tested specification was then used to generate corresponding machine-readable state transition automata and integrated into the *Generalized Dialogue Manager* of MIGSEP.

4.3 The Elderly-friendly Design Elements in MIGSEP

According to the design guidelines in the previous section, a set of elderly-centered design elements were implemented in MIGSEP. Specifically, the most essential elements are listed below:

- **Visual perception**: simple and clear layout was constructed without overlapping items; 12-14 sized sans-serif fonts were chosen for all displayed texts. Simple and high contrast colours without fancy visual effects were used and placed aside; regularly shaped rectangles and circles were selected, enabling comfortabe perception and easy recognition.
- **Hearing ability**: both text and acoustic output are provided as system responses. Styles, vocabulary, structures of the sentences have been intensively revised. A low-pitched yet vigorous male voice is chosen for the synthesis.
- **Motor functions**: regularly shaped, sufficiently sized and well separated interface elements were designed for easy access. Clicking was decided to be the only action to avoid otherwise frequently occurring errors caused by decline of motor and attentional functions. "Start" was provided as the only way of orientating oneself to avoid confusion.
- **Attention or Concentration**: fancy irrelevant images or decorations were avoided. Unified font, colors, sizes of interface elements were used for the entire interface. Simple animation notifying changes were constructed, giving sufficiently clear yet not distracting feedback to the user.
- **Memory abilities**: all items are used with relevant keywords. The number of displayed items is restricted to no more than three, considering the maximum capacity of short term memory, the accessible size as well as the readable amount of information of the interaction items on a table PC. Logically well-structured and sequentially presented items were intensively revised to assist orientation during interaction. Context sensitive clues are given with selected colors.
- **Intellectual ability**: consistent layout, colours and interaction styles are used. Changes on the interface happen only on data level.

4.4 Interaction with MIGSEP in Hospital Environments

We have implemented a MIGSEP system and set its application domain to hospital environments. Figure 12 shows a user interacting with it via speech modality.

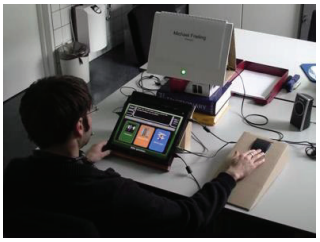


Figure 12: a user is interacting with MIGSEP.

Figure 13 shows a sample dialogue between the MIGSEP system and a user who would like to be guided to the cardiology department, to a doctor named Wolf.

S: <shows 3 cards, green for persons, yellow for rooms and blue for departments>
hello, where do you want to go?
U: I want to go to Wolf in the cardiology department.
S: <shows a blue bar texted with cardilogy on the top, and 2 green cards, with photos, titles and names in the middle>
I found two persons working in the cardiology department with the name wolf, to which do you want to go?
U: to Dr. Wolf please.
S: <enlarges the second green card>
Do you want to go to Mrs. Dr. Diana Wolf?
U: Yes please.
S: <resizes the second green card and shows the 3 cards in the start menu again>
OK, I have saved this goal. Where else do you want to go?

Figure 13: example of a dialogue with MIGSEP.

5 THE EXPERIMENTAL STUDY

To evaluate how well elderly is assisted by MIGSEP system, an experimental study was conducted.

5.1 Participants

Eighteen elderly persons (m/f: 11/7, mean age of 70.9, standard deviation (SD)=3.0), all German native speakers, took part in the study. They all had the mini-mental state examination (MMSE), which is a screening test to measure cognitive mental status (Folstein, Folstein and Mchugh, 1975). A test value between 28 and 30 indicates normal cognitive functioning, therefore, our participants showing 28.3 (SD=.86) were in the normal range.

5.2 Stimuli and Apparatus

As shown in figure 12, visual stimuli were given by the green lamp and the graphical user interface on the screen of a portable tablet PC; audio stimuli as complementary feedbacks were also generated by the MIGSEP system and presented via two loudspeakers at a well-perceivable volume. All tasks were given as keywords on the pages of a calendar-like system. The only input possibility was the spoken language instructions, activated if the button was being pressed and the green lamp was on.

The same data set contains virtual information about personnel, rooms and departments in a common hospital, was used in the experiment.

During the experiment each participant was accompanied by only one investigator, who gave the introduction and well-defined instructions at the beginning, and provided help if necessary (which was very rare the case).

An automatic internal logger of the MIGSEP system was used to collect the real-time data, while the windows standard audio recorder program kept track of the whole dialogic interaction process.

A questionnaire focusing on the user satisfaction was designed. It includes questions of seven categories: system behaviour, speech output, textual output, interface presentation, task performing, user-friendliness and user perspective. The questionnaire was completed by each participant by a five point Likert scale, where one represents the lowest appropriateness and five the highest.

5.3 Procedure

Each participant had to undergo four phases:

- **Introduction**: a brief introduction was given to the participants.
- **Learning**: they were instructed how to interact with the MIGSEP system using the button device and spoken natural language. After they made no more mistakes using the button device, a further introduction was given to the verbal and graphical feedbacks the system provides. Then they were asked to perform one or two sample tasks to gather more practical experiences with the system.
- **Testing**: Each participant had to perform eleven tasks, each of which contains incomplete yet sufficient information about a destination the participant should select. Each task was ended, if the goal was selected, or the participant gave up trying after six minutes.
- **Evaluation**: After all tasks were run through, each participant was asked to fill in the questionnaire for evaluation.

5.4 Questions and Methods

Altogether, there are three important questions to be focused and answered by the experiment:

- ”Can elderly use the MIGSEP system to complete the tasks?”
  A standard measurement method Kappa coefficient is used to assess the successfulness of the interaction between the participants and the system.
- “Can elderly persons handle the tasks with MIGSEP efficiently?”
  This shall be answered by the automatically logged data of every single interaction.
- “Do elderly find it comfortable to interact with MIGSEP?”
  This should be reflected in the results of the evaluation questionnaires.

6 RESULTS

6.1 Effectiveness of MIGSEP

To answer the first question, i.e., how well the MIGSEP system assists elderly persons to perform tasks, we used Kappa coefficient, which is a well-accepted method for measuring effectiveness of interaction (Walker, et al., 1997).

In order to apply this method, we needed to define the attribute value matrix (AVM), which had to contain all information that has to be exchanged between MIGSEP and the participants. E.g. table 1 shows the AVM for the task: ”Drive to a person named Michael Frieling.”, where the expected values of this task are also presented.

Table 1: An example AVM for the task “drive to a person name Michael Frieling”.

Attribute	Expected value
FN	Michael
LN	Frieling
G	Male

By combining the actual data recorded during the experiment with the expected attribute values in the AVMs, we can construct the confusion matrices for all tasks. E.g., table 2 shows the confusion matrix for the task ”drive to a person named Michael Frieling”, where ”M” and ”N” denote whether the actual data match with the expected attribute values in the AVMs. E.g. one participant selected a person with wrong first and last names.

Table 2: The confusion matrix for the task “drive to a person named Michael Frieling”.

	FN		LN		G		
Data	M	N	M	N	M	N	sum
FN	17	1					18
LN			17	1			18
G					18		18

Given one confusion matrix, the Kappa coefficient can be calculated with

$\kappa = \frac{P(A)-P(E)}{1-P(E)}$, (Walker, et al., 1997)

In our experiment,

$P(A) = \frac{\sum_{i=1}^{n} M(i,\ M)}{T}$

is the proportion of times that the actual data agree with the attribute values, and

$P(E) = \sum_{i=1}^{n} (\frac{M(i)}{T})^2$

is the proportion of times that the actual data are expected to be agreed by chance, where M(i, M) is the value of the matched cell of row i, M(i) the sum of the cells of row i, and T the sum of all cells.

Therefore, we summarized the results of all the tasks and constructed one confusion matrix for all the data, and got that, $P(A) = 0.961$ and kappa coefficient $\kappa = 0.955$, which suggests a highly successful degree of interaction between the MIGSEP system and the participants.

6.2 Efficiency of MIGSEP

Regarding the efficiency of MIGSEP, quantitative data automatically logged during the experiments are summarized in table 3, with respect to user turns, system turns, ASR failed times (the frequency of the Automatic Speech Recognizer failing getting a parsable sentence), ASR error times (the frequency of the ASR wrongly recognizing utterances), user turns without ASR (user turns without being affected by the ASR related failures) and the elapsed time for each participant and each task.

Table 3: Quantitative results calculated based on the recorded data concerning efficiency.

	Average	Standard deviation
User turns	4.1	1.8
Sys turns	4.0	1.7
ASR failed times	1.2	0.8
ASR error times	1.0	1.2
User turns without ASR	1.9	0.4
Elapsed time	61.0	23.6

From a dialogue system's points of view, a very good overall performance efficiency is shown by averagely 4.1 user turns and 3.9 system turns per task for each participant, as the average basic turn numbers, which can be inferred by the shortest solution regarding the number of slots for each task to be filled, are 3 user turns and 3 system turns. In addition, if the ASR related failures and errors are excluded, the user turns would be only 1.9. This shows that almost each task was completed by each participant with only one complicated sentence. Furthermore, the user turns without ASR, which is lower than the theoretically minimum 2 user turns, even implied that with slightly wrong recognized sentence, the MIGSEP system was still able to find a solution to help elderly persons to complete tasks.

On the other hand, the elapsed time for each task and each participant is considered as satisfying, with averagely 61.0 second for minimal 6 interaction paces (3 user turns +3 system turns), including the relatively long spoken utterance either by the system or the elderly participants. However, the standard deviation of 23.6 is a bit high, since two participants needed much longer time than the others. They encountered many problems with the automatic speech recognizer, which indicates the necessity for further analysis and improvement of the ASR.

6.3 User Satisfaction

Table 5: The assessment of subjective user satisfaction.

	Mean	Standard deviation
System behaviour	3.7	0.8
Speech output	4.5	0.5
Textual output	4.7	0.5
Interface presentation	4.6	0.4
Task performing	4.3	0.4
User-friendliness	4.6	0.4
User perspective	3.9	0.8
Overall	4.3	0.4

Overall, it shows a very good user satisfaction with the averagely score of 4.3 out of 5. Specifically, the speech and textual outputs are considered appropriately constructed with the score of 4.5 and 4.7; the interface is intuitive and easy to understand with the score of 4.6; the process to perform the task is quite feasible with the score of 4.3; and the system is rather user-friendly with the score of 4.6 out of 5.

However, the scores of system behaviour and user perspective were a bit lower than the others. It is mainly due to the problem of the automatic speech recognizer, which could trigger unexpected system responses, and therefore make the future use from the user perspective less attractive.

7 CONCLUSIONS AND FUTURE WORK

This paper presented our work on multimodal interaction for elderly persons from three essential perspectives: the modelling and development of multimodal interaction using a tool-supported, formally tractable and extensible unified dialogue modelling approach; the design and implementation of a multimodal interactive system according to a number of elderly-friendly guidelines regarding the basic design principles of conventional interactive interfaces and ageing centered characteristics. The multimodal interactive system was evaluated with eighteen elderly persons. The evaluation showed high effectiveness, high efficiency and a high satisfaction of the user with our system. These findings provide us with further evidence for our proposed guidelines, approaches and frameworks on system design and implementation.

The presented work served as part of a developmental process towards building an effective, efficient, adaptive and robust multimodal interactive framework for the elderly. Further study focussing on speech and touch screen combined modalities is being conducted. Moreover, corpus-based supervised and reinforcement learning techniques will be applied to improve the current dialogue model and gain more flexible interaction to compensate for the insufficient reliability of automatic speech recognizers. Our future research will continue with combining and experimenting emerging technologies in addition to speech, touch screen and visual modalities. Special attentions are also being paid to learning-based discourse modelling and management in advanced multimodal interactive systems for elderly persons.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG) through the Collaborative Research Center SFB/TR8, the department of Medical Psychology and Medical Sociology and the department of Neurology of the University Medical Center Göttingen, and the German Research Centre for Artificial Intelligence.

REFERENCES

Alston, P.W., 2000, *Illocutionary acts and sentence meaning.* Cornell University Press.

Birdi, K., Pennington, J., Zapf, D., 1997. Aging and errors in computer based work: an observational field study. In *Journal of Occupational and Organizational Psychology.* pp. 35-74.

Bolt, R.A., 1980. Put-That-There: Voice and Gesture at the Graphics Interface. In *Proceedings of the 7th International Conference on Computer Graphics and Interactive Techniques.* Seattle, USA, pp. 262-270.

Broadfoot, P., Roscoe, B., 2000. Tutorial on FDR and Its Applications. In *K. Havelund, J. Penix and W. Visser (eds.), SPIN model checking and software verification.* Springer-Verlag, London, UK, Volume 1885, pp. 322.

Charness, N., Bosman, E., 1990. Human Factors and Design. In *J.E. Birren and K.W. Schaie, (eds.), Handbook of the Psychology of Aging.* Academic Press, Volume 3, pp. 446-463.

Craik, F., Jennings, J., 1992. Human memory. In *F. Craik and T.A. Salthouse, (eds.), The Handbook of Aging and Cognition.* Erlbaum, pp. 51-110.

Folstein, M., Folstein, S., Mchugh, P., 1975. "mini-mental state", a practical method for grading the cognitive state of patients for clinician. In *Journal of Psychiatric Research.* Volume 12, 3, pp. 189-198.

Fozard, J.L., 1990. Vision and hearing in aging. In *J. Birren, R. Sloane and G.D. Cohen (eds), Handbook of Metal Health and Aging.* Academic Press, Volume 3, pp. 18-21.

Gawthorn, D., 2000. Possible implications of ageing for interface designer. In *Interacting with Computers.* pp. 507-528.

Hall, A., Chapman, R., 2002. Correctness by construction: Developing a commercial secure system. In *IEEE Software.* Vol. 19, 1, pp. 18-25.

Holzinger, A., Mukasa, K.S., Nischelwitzer, A.K., 2008. Introduction to the special thematic session: Human-computer interaction and usability for elderly. In *Proceedings of the 11th International Conference on Computers Help People with Special Needs.* Springer Verlag, Berlin, Germany, pp. 18-21.

Hoyer, W.J., Rybash, J.M., 1992. Age and visual field differences in computing visual spatial relations. In *Psychology and Aging 7.* pp. 339-342.

Jaimes, A., Sebe N., 2007. Multimodal human-computer interaction: A survey. In *Computational Vision and Image Understanding. Elsevier Science Inc.,* New York, USA, pp. 116-134.

Jian, C., Scharfmeister, F., Rachuy, C., Sasse, N., Shi, H., Schmidt, H., Steinbüchel-Rheinwll, N. v., 2011. Towards Effective, Efficient and Elderly-friendly Multimodal Interaction. In *PETRA 2011: Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments. ACM,* New York, USA.

Kline, D.W., Scialfa, C.T., 1996. Sensory and Perceptual Functioning: basic research and human factors implications. In *A.D. Fisk and W.A. Rogers. (eds.), Handbook of Human Factors and the Older Adult,* Academic Press.

Kotary, L., Hoyer, W.J., 1995. Age and the ability to inhibit distractor information in visual selective attention. In *Experimental Aging Research.* Volume 21, Issue 2.

Mackay, D., Abrams, L., 1996. Language, memory and aging. In *J.E. Birren and K.W.Schaie (eds), Handbook of the psychology of Aging*. Academic Press, Volume 4, pp. 251-265.

McDowd, J.M., Craik, F. 1988. Effects of aging and task difficulty on divided attention performance. In *Journal of Experimental Psychology: Human Perception and Performance 14*. pp. 267-280.

Moeller, S., Goedde, F., Wolters, M., 2008. Corpus analysis of spoken smart-home interactions with older users. In N. Calzolari, K.Choukri, B. Maegaard, J. Mariani, J. Odjik, S. Piperidis, and D. Tapias, (eds.), Proceedings of the Sixth International Conference on Language Resources Association. ELRA.

Morris, J.M., 1994. User interface design for older adults. In *Interacting with Computers.* Vol. 6, 4, pp. 373-393.

Oviatt, S. T., 1999. Ten myths of multimodal interaction. In *Communications of the ACM.* ACM New York, USA, Vol. 42, No. 11, pp. 74-81.

Roscoe, A.W., 1997. *The Theory and Practice of Concurrency,* Prentice Hall.

Ross, J.R., Bateman, J., Shi, H., 2005. Using Generalized Dialogue Models to Constrain Information State Based Dialogue Systems. In *Symposium on Dialogue Modelling and Generation.*

Salthouse, T.A., 1994. The aging of working memory. In *Neuropsychology 8,* pp. 535-543.

Schieber, F., 1992. Aging and the senses. In *J.E. Birren, R.B. Sloane, and G.D. Cohen, (eds.) Handbook of Mental Health and Aging,* Academic Press, Volume 2.

Shaie, K.W., 1996. Intellectual development in adulthood. In *J.E. Birren and K.W. Shaie, (eds.), Handbook of the psychology of aging.* Academic Press, Volume 4.

Shi, H., Bateman, J., 2005. Developing human-robot dialogue management formally. In Proceedings of Symposium on Dialogue Modelling and Generation. Amsterdam, Netherlands.

Sitter, S., Stein, A., 1992. Modelling the illocutionary aspects of information-seeking dialogues. In *Journal of Information Processing and Management.* Elsevier, Volume 28, issue 2, pp. 165-180.

Traum, D., Larsson, S., 2003. The information state approach to dialogue management. In *J.v. Kuppevelt and R. Smith (eds.), Current and New Directions in Discourse and Dialogue.* Kluwer, pp. 325-354.

Walkder, N., Philbin, D.A., Fisk, A.D., 1997. Age-related differences in movement control: adjust submovement structure to optimize performance. In *Journal of Gerontology: Psychological Sciences 52B,* pp. 40-52.

Walker, M.A., Litman, D.J., Kamm, C.A., Kamm, A.A., Abella, A., 1997. Paradise: a framework for evaluating spoken dialogue agents. In *Proceedings of the eighth conference on European chapter of Association for computational Linguistics,* NJ, USA, pp. 271-280.

Touch and Speech: Multimodal Interaction for Elderly Persons

Cui Jian[1], Hui Shi[1], Frank Schafmeister[2], Carsten Rachuy[1], Nadine Sasse[2], Holger Schmidt[3], Volker Hoemberg[4], and Nicole von Steinbüchel[2]

[1]SFB/TR8 Spatial Cognition, Universität Bremen, Germany
{ken, shi, rachuy}@informatik.uni-bremen.de
[2]Medical Psychology and Medical Sociology, University Medical Center Göttingen, Germany
{frank-schafmeister, n.sasse, nvsteinbuechel}@med.uni-goettingen.de
[3]Neurology, University Medical Center Göttingen, Germany
h.schmidt@med.uni-goettingen.de
[4]SRH Health Centre Bad Wimpfen
mueller-hoemberg@t-online.de

Abstract. This paper reports our work on the development and evaluation of a multimodal interactive guidance system for navigating elderly persons in hospital environments. A list of design guidelines has been proposed and implemented in our system, addressing the needs of designing a multimodal interfaces for elderly persons. Meanwhile, the central component of an interactive system, the dialogue manager, has been developed according to a unified dialogue modelling method, which combines the conventional recursive transition network based generalized dialogue models and the classic agent-based dialogue theory, and supported by a formal language based development toolkit. In order to evaluate the minutely developed multimodal interactive system, the touch and speech input modalities of the current system were evaluated by an elaborated experimental study with altogether 31 elderly. The overall positive results on the effectiveness, efficiency and user satisfaction of both modalities confirm our proposed guidelines, approaches and frameworks on interactive system development. Despite the slightly different results, there is no significant evidence for one preferred modality. Thus, further study of their combination is considered necessary.

Keywords: Multimodal interaction, elderly-centered system design, human-computer interaction, spoken dialogue systems, formal methods

1 Introduction

Multimodal interfaces is gaining more and more importance for its promising possibility to achieve a significantly more effective and efficient human computer interaction (cf. [1]), they also increase users' satisfaction and provide a more natural and intuitive way of interaction (cf. [2]). Meanwhile, due to the demographic development

adfa, p. 1, 2011.

towards increasingly more elderly persons, there rises a growing research focus on the multimodal communication technology, which aims at enhancing the quality of interaction by taking age-related decline into account (cf. [3]).

Elderly people often suffer from decline of sensory, perceptual, motor and cognitive abilities. Considering these facts we first present a list of elaborated design guidelines regarding basic design principles of conventional interactive systems and the most common elderly-centered characteristics. Meanwhile, in order to achieve a flexible and context-sensitive, yet formally tractable and controllable interaction, we designed a unified dialogue modelling approach, which combines a finite state based generalized dialogue model and the classic agent based dialogue model, and implemented this by a formal language based development framework. According to the proposed design guidelines and the unified dialogue modelling approach, an interactive guidance system was especially designed and developed for the elderly. To evaluate the touch input and natural spoken language modalities with respect to their feasibility and acceptance by elderly persons, an empirical study was conducted with 31 older participants. The general framework PARADISE [4] has been applied in our evaluation process. The study also aimed at the evaluation of the multimodal interactive guidance system as a whole, while regarding the essential criteria of the following aspects for interaction: the effectiveness of task success, the efficiency of executing tasks and the user satisfaction with the system.

The following text is organized as follows: section 2 presents the proposed general guidelines for designing multimodal interactive system for elderly persons; section 3 introduces the unified dialogue modelling approach which combines the classic agent based approach and the recursive transition network based theory for building the discourse management of the multimodal interaction; section 4 then describes the multimodal interactive guidance system, which is developed based on the unified dialogue modelling approach and the presented set of design guidelines; section 5 describes the experimental study, and the results are analyzed and discussed in section 6. Finally, section 7 concludes and gives an outline of the future work.

2 Design Guidelines of Multimodal Interactive Systems for Elderly Persons

[5] indicated that the decline of elderly persons should be considered while designing interactive systems for the elderly. Therefore, we defined a set of design guidelines for multimodal interaction with respect to the decline of seven very important abilities. They are implemented and integrated into our multimodal interactive guidance system and tested in an empirical pilot study. The results are described in [6] and the improved guidelines are presented as follows:

2.1 Visual Perception

Visual perception declines for most people with age (cf. [7]) in different ways: many people find it more difficult to focus on objects up close and to see fine details; the

size of the visual field is decreasing and the peripheral vision is successively declining. Rich colors and complex shapes are making perception difficult. Rapidly moving objects are either causing too much distraction, or becoming less noticeable. To cope with these impairments, the following guidelines should be taken into account:

- Layouts of the user interface should be devised as simple and clear as possible, with few (if any) or no overlapping items.
- All texts should be large enough to be readable on the communicating interfaces.
- Strong contrast should be used with as few colors as possible; this also applies to simple and easily recognizable shape designs.
- Unnecessary and irrelevant visual effects should be avoided.

2.2 Speech Ability

Elderly persons need more time to produce complex words or longer sentences, probably due to reduced motor control of tongue and lips (cf. [8]). Furthermore, speech-related adaptation is necessary to improve the interaction quality to a sufficient level (cf. [9]). Based on these, the following aspects should be taken into account:

- Special acoustic models for the elderly should be used for speech recognizer.
- Vocabulary should be built with more definite articles, auxiliaries, first person pronouns and lexical items related to social interaction. Texts should be as simple as possible.
- Dialogue strategies should be able to cope with elderly specific needs such as repeating, helping and social interaction, etc.

2.3 Auditory Perception

Hearing ability declines at least to 75% after 75 year olds (cf. [10]). High pitched sounds are increasingly lost, as well as long and complex sentences becoming difficult to follow (cf. [11]). Therefore special attention should be paid to the following:

- Text displays can help when information is mis- or not heard, which should not provide conflicting information.
- Synthesized texts should be intensively revised regarding style, vocabulary, length and sentence structures suitable for elderly.
- Low pitched voices are more acceptable for speech synthesis, e.g., female voices are less preferred than male ones.

2.4 Motor Ability

Computer mice are unsuitable for many elderly due to the lack of good hand-eye coordination and decline of fine motor abilities (cf. [12]). Positioning the cursor is difficult if the target is too small or too irregular to locate, and fine movements are harder to control (cf. [13]).Thus, the following procedures are suggested:

- Direct interaction is recommended, e.g., touch screen.
- All GUI items should be accessibly shaped, sized and well spaced from each other.
- Simple movements are recommended, such as clicking instead of dragging.
- Text input should be avoided or replaced with other simpler input actions.
- An undo function is needed to correct errors.
- Simultaneous multimodal input such as the combination of speech and other input should be avoided or replaced.

2.5 Attention and Concentration

Elderly persons are more easily distracted by details or noise (cf. [14]). They show great difficulty maintaining divided attention, where attention must be paid to more than one aspect at a time (cf. [15]). Therefore, the following points are suggested:

- Only relevant images should be used.
- Items should not be displayed simultaneously.
- Unified or similar fonts, colors and sizes of displayed texts are recommended.
- Changes on the user interface should be emphasized in an obvious way.

2.6 Memory Functionalities

Different memory functions decline at different degrees during ageing. Short term memory holds fewer items while ageing and declines earlier; also more time is needed to process information (cf. [16]). Working memory also becomes less efficient (cf. [17]). Semantic information is believed to be preserved in long term memory for a longer period(cf. [18]). To compensate the decline of the different memory functions, the following points are suggested:

- Pure image items should be avoided or placed near relevant key words.
- Presented items in a sequence should not exceed five, the average maximum capacity of short term memory of elderly persons.
- Information should be categorized to assist storage into long term memory.
- Context sensitive information is necessary to facilitate working memory activities.

2.7 Intellectual Ability

Fluid intelligence does decline with ageing, while, crystallized intelligence does not or to a less extent (cf. [19]); it can assist elderly people to perform better in a stable well-known interface environment. Thus, we suggest assuring the following points:

- Unified interface layout, where changes should only happen on data level.
- Semantically intuitive structure, where users should not be too surprised while traversing the interaction levels.
- Consistent interaction style facilitates learning and assist elderly to master interface use.

3 A Formal Unified Dialogue Modelling Approach

As a typical recursive network based approach, generalized dialogue models were developed by constructing dialogue structures at the illocutionary level (cf. [20]). However, it is criticized for its inflexibility of dealing with dynamic information exchange. Meanwhile, information state update based theories were deemed the most successful foundation of agent based dialogue approaches (cf. [21]), which provides a powerful mechanism to handle dynamic information and gains a context sensitive dialogue management. Nevertheless, such models are usually very difficult to manage and extend (cf. [22]).

Thus, a unified dialogue modelling approach was developed. It combines the generalized dialogue models with information state updated based theories. This approach is supported by a formal development toolkit, which is used to implement an effective, flexible, yet formally controllable dialogue management.

3.1 A Unified Dialogue Modelling Approach

Generalized dialogue models can be constructed with the recursive transition networks (RTN). They abstract dialogue models by describing illocutionary acts without reference to direct surface indicators (cf. [23]). Fig. 1 (left) shows a simple generalized dialogue model as a recursive transition network diagram. It is initiated with an assertion from a person A, responded by B with three actions: accept, agree or reject.

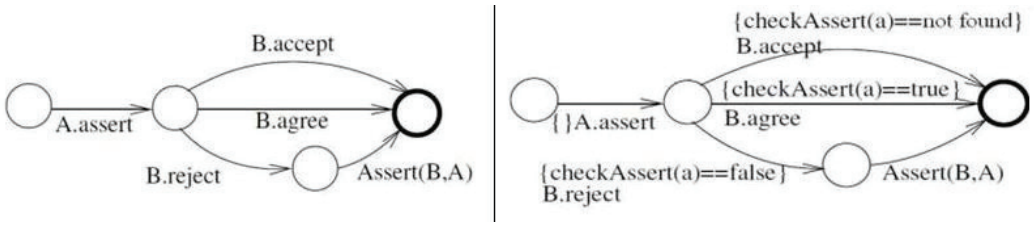


Fig. 1. A generalized dialogue model as a simple recursive transition network (RTN) (left) & a generalized dialogue model as a simple deterministic RTN with conditional transitions (right)

The generalized dialogue model above is a none-deterministic model. To build a feasible interaction model, deterministic behavior should be assured for the interaction flow. Thus, conditional transitions are introduced to improve the above dialogue model (cf. Fig. 1 right). Let *checkAssert* be a method to check whether an assertion holds with B's knowledge and *a* an assertion given by A, if the assertion holds, B can agree with it; otherwise, B rejects it and initiates further discussion; if the assertion is not known by B, then B accepts it. Such conditional transitions can only be activated if the relevant condition is fulfilled. We call it the conditional RTN.

Although the conditional RTN based generalized dialogue model defines a deterministic illocutionary structure, it does not provide the mechanism to integrate discourse information. Thus, information state based theory was integrated into our unified dialogue model by eliminating some typical elements, e.g. AGENDA for planning the next dialogue moves, because such information is already captured by the

generalized dialogue model; furthermore it complements illocutionary structure with update rules, which is associated with the information state of current context, and can update the information state respectively if necessary. As a result, a unified dialogue model is constructed as shown in Fig. 2 (left). Four update rules are added, so that the discourse context can always be considered and updated; e.g. the update rule ACCEPT adds a new assertion *a* into B's belief and refer it as known from then on.

Finally, we define a unified dialogue model as a deterministic recursive transition network built at the illocutionary level of interaction processes; its transitions can only be triggered by fulfilled conditions concerning the information state, and with the consequences of information state update according to a set of update rules.

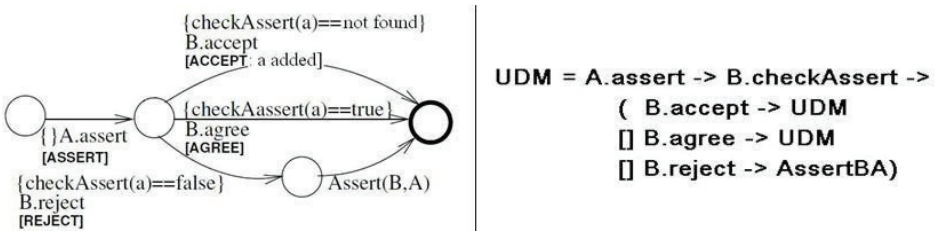


Fig. 2. A simple unified dialogue model and its CSP specification

3.2 A Formal Language Based Development Toolkit for Dialogue Modelling

Deterministic recursive transition networks can be illustrated as a typical finite state transition diagram (cf. Fig. 2 left), which provides the possibility of specifying the described illocutionary structure with mathematically well-founded formal methods, e.g., with Communicating Sequential Processes (CSP) in the formal methods community of computer science.

CSP can not only be used to specify finite state automata structured patterns with abstract, yet highly readable and easily maintainable logic formalization (cf. [24]), but it is also supported by well-established model checkers to verify the concurrent aspects and increase the tractability (cf. [25]). Thus, CSP is used to specify and verify the unified dialogue models (cf. the example in Fig. 2 (right)).

To support the development of unified dialogue models within interactive systems, we provided the Formal Dialogue Development Toolkit (FormDia cf. Fig. 3).

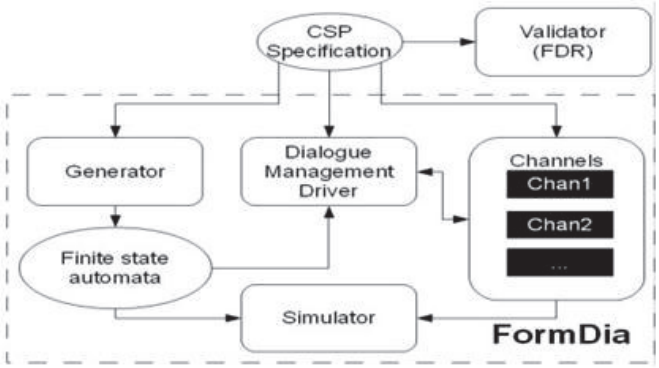


Fig. 3. the Structure of the FormDia Toolkit (cf. [26])

To develop the unified dialogue model based discourse management, FormDia toolkit can be used according to the following steps:

- Validation: the CSP specified structure of a unified dialogue model can be validated by using Failures-Divergence Refinement tool, (FDR (cf. [27])), which is a model checking tool for validating and verifying concurrency of state automata.
- Generation: according to the given CSP specification, finite state automata can then be generated by the FormDia Generator.
- Channels Definition: channels between the dialogue management and application/domain specific components can be defined. These channels are at first black boxes, which will be filled with deterministic behavior of concrete components.
- Simulation: with the generated finite state automata and the communication channels, dialogues scenarios are simulated via a graphical interface, which visualizes dialogue states as a directed graph and provides a set of utilities to trigger events and the dialogue state update for testing and verification.
- Integration: after the dialogue model is validated and verified, it can be integrated into a practical interactive dialogue system via a dialogue management driver.

The FormDia toolkit shows a promising way for developing formally tractable and extensible interaction. It enables an intuitive design of dialogue models with formal language, automatic validation of related functional properties, and it also provides an easy simulation, verification for the specified dialogue models, and the straightforward integration within a practical interactive system. In addition, with the unified dialogue model, FormDia toolkit can even be used in multimodal interactive system.

4 Multimodal Interactive Guidance System for Elderly Persons

The Multimodal Interactive Guidance System for Elderly Persons (MIGSEP) was developed for elderly or handicapped persons to navigate through public spaces. MIGSEP runs on a portable touch screen tablet PC. It serves as the interactive media designed for an autonomous intelligent electronic wheelchair that can automatically carry its users to desired locations within complex environments.

4.1 System Architecture

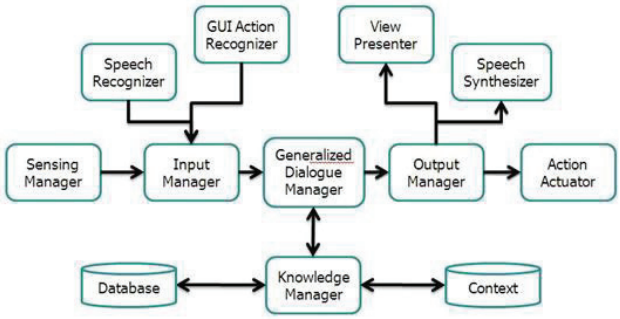


Fig. 4. The architecture of MIGSEP

The architecture of MIGSEP is illustrated in Fig. 4. A Generalized Dialogue Manager is developed using the unified dialogue modelling approach. It functions as the central processing unit and enables a formally controllable and extensible, meanwhile context-sensitive multimodal interaction. An Input Manager receives and interprets all incoming messages from GUI Action Recognizer for GUI inputs, Speech Recognizer for natural language understanding and Sensing Manager for other sensor data. An Output Manager on the other hand, handles all outgoing commands and distributes them to View Presenter for visual feedbacks, Speech Synthesizer to generate natural language responses and Action Actuator to perform necessary motor actions. Knowledge Manager uses Database to keep the static data of certain environments and Context to process the dynamic information exchanged with users during the interaction.

Although the essential components of MIGSEP are closely connected with each other via predefined XML-based communication mechanism, each of them is treated as an open black box and can be implemented or extended for specific use, without affecting other MIGSEP components. It provides a general platform for both theoretical researches and empirical studies on multimodal interaction.

4.2 The Unified Dialogue Model in MIGSEP

The current unified dialogue model (UDM) consists of four extended state transition diagrams.

Each interaction is initiated with the diagram Dialogue(S, U) (cf. Fig. 5 (left)), by the initialization of the system's start state and a greeting-like request.

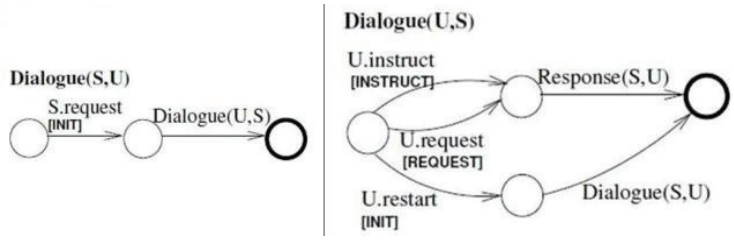


Fig. 5. The Initiate diagram and the transition diagram triggered by user

The dialogue continues with user's instruction to a certain location, request for a certain information or restart action, leading to the system's further response or dialogue restart, respectively, as well as updating the information state with the attached update rules (cf. *Dialogue(U, S)* in Fig. 5 (right)).

After receiving user's input, the system tries to generate an appropriate response according to its current knowledge base and information state (cf. Response(S, U) in Fig. 6 (left)). This can be informing the user with requested data, rejecting an unacceptable request with or without certain reasons, providing choices for multiple options, or asking for further confirmation of taking a critical action, each of which triggers transitions to different diagrams.

Finally, the user can accept or reject the system's response, or even ignore it by simply providing new instructions or requests, triggering further state transitions as well as information state updates (cf. Response(U, S) in Fig. 6 (right)).

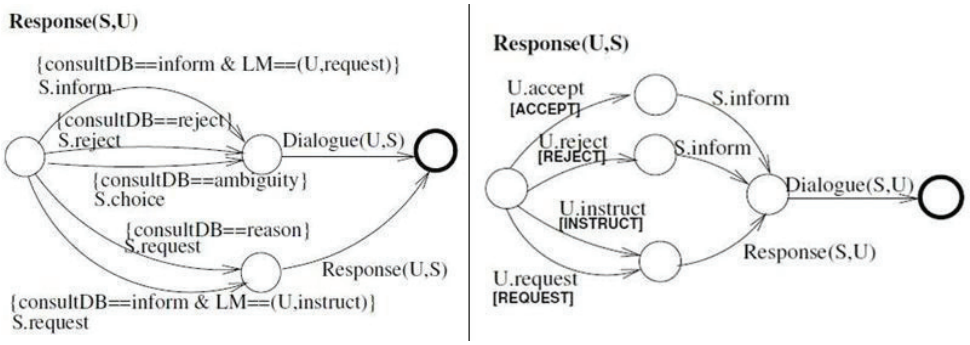


Fig. 6. The model for system's response and the model for user's response

Using the FormDia toolkit, the UDM was developed as CSP specifications, and its functional properties have been validated and verified via FDR, as well as its conceptual interaction process using FormDia simulator. The tested specification was then used to generate corresponding machine-readable state transition automata and integrated into the Generalized Dialogue Manager of MIGSEP.

4.3 The Elderly-friendly Design Elements in MIGSEP

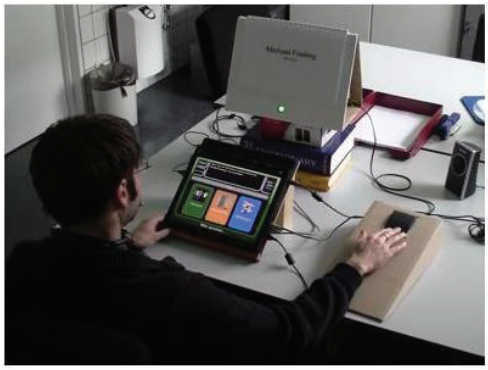
According to the design guidelines in the previous section, a set of elderly-centered design elements were implemented in MIGSEP. Specifically, the most essential elements are listed below:

- Visual perception: simple and clear layout was constructed without overlapping items; 12-14 sized sans-serif fonts were chosen for all displayed texts. Simple and high contrast colors without fancy visual effects were used and placed aside; regularly shaped rectangles and circles were selected, enabling comfortable perception and easy recognition.
- Auditory perception: both text and acoustic output are provided as system responses. Styles, vocabulary, structures of the sentences have been intensively revised. A low-pitched yet vigorous male voice is chosen for the synthesis.
- Motor functions: regularly shaped, sufficiently sized and well separated interface elements were designed for easy access. Clicking was decided to be the only action to avoid otherwise frequently occurring errors caused by decline of motor function. "Start" was provided as the only way of orientating oneself to avoid confusion.
- Attention/Concentration: fancy irrelevant images or decorations were avoided. Unified font, colors, sizes of interface elements were used for the entire interface. Simple animation notifying changes were constructed, giving sufficiently clear yet not distracting feedback to the user.
- Memory functionalities: all items are used with relevant keywords. The number of displayed items is restricted to no more than three, considering the maximum capacity of short term memory, the accessible size as well as the readable amount of information of the interaction items on a table PC. Logically well-structured and

sequentially presented items were intensively revised to assist orientation during interaction. Context sensitive clues are given with selected colors.
- Intellectual ability: consistent layout, colors and interaction styles are used. Changes on the interface happen only on data level.

4.4 Interaction with MIGSEP in a Hospital Environment

We have implemented a MIGSEP system and set its application domain to hospital environments. Fig. 7 shows a user interacting with it via speech modality and a sample dialogue between the MIGSEP system and a user who would like to be guided to the cardiology department, to a doctor named Wolf.



S: <shows 3 cards, green for persons, yellow for rooms and blue for departments>
hello, where do you want to go?
U: I want to go to Wolf in the cardiology department.
S: <shows a blue bar texted with cardilogy on the top, and 2 green cards, with photos, titles and names in the middle>
I found two persons working in the cardiology department with the name wolf, to which do you want to go?
U: to Dr. Wolf please.
S: <enlarges the second green card>
Do you want to go to Mrs. Dr. Diana Wolf?
U: Yes please.
S: <resizes the second green card and shows the 3 cards in the start menu again>
OK, I have saved this goal. Where else do you want to go?

Fig. 7. A user is interacting with MIGSEP and an Example of a dialogue with MIGSEP

5 The Experimental Study

To evaluate how well an elderly person is assisted by MIGSEP system, an experimental study concerning speech and touch input modalities was conducted.

5.1 Participants

Altogether 31 elderly persons (m/f: 18/13, mean age of 70.7, standard deviation 3.1), all German native speakers, took part in the study, in which 15 participants were using the speech input and 16 were using touch input. They all finished the mini-mental state examination (MMSE), which is a screening test to measure cognitive mental status (cf. [28]). A test value between 28 and 30 indicates slight decline yet sufficiently normal cognitive functioning, therefore, our participants showing 29.0 averagely (std.=.84) were in the acceptable range.

5.2 Stimuli and Apparatus

Except the variation of the input possibilities between touch and speech, where the touch modality was supported by a touchable screen of a laptop and the spoken language instructions were only activated if the button was being pressed and the green

lamp was on (cf. Fig. 7), all other stimuli are the same for both modalities, e.g., visual stimuli were given by a green lamp and a graphical user interface; audio stimuli as complementary feedbacks were also generated by the MIGSEP system and presented via two loudspeakers at a well-perceivable volume. All tasks were given as keywords on the pages of a calendar-like system.

The same data set contains virtual information about personnel, rooms and departments in a common hospital, was used for both modalities throughout the experiment.

During the experiment each participant was accompanied by only one investigator, who gave the introduction and well-defined instructions at the beginning, and provided help if necessary.

An automatic internal logger of the MIGSEP system was used to collect the real-time data, while the windows standard audio recorder program kept track of the whole dialogic interaction process.

A questionnaire, which is focusing on the user satisfaction and includes questions of seven categories: system behavior, speech output, textual output, interface presentation, task performing, user-friendliness and user perspective, was filled in by each participant via a five point Likert scale based grading system.

5.3 Procedure

Each participant (both touch and speech) had to undergo four phases:

- Introduction: a brief introduction was given to the participants.
- Standardized learning phase: they were instructed how to interact with the MIGSEP system, either using the touchable screen or using the button device and spoken natural language. After they made no more mistakes with the assigned input modality, a further introduction was given to the verbal and graphical feedbacks the system provides. Then they were asked to perform one task to gather more practical experiences.
- Testing: Each participant had to perform 11 tasks, each of which contains incomplete yet sufficient information about a destination. Each task was ended, if the goal was selected, or the participant gave up trying after six minutes.
- Evaluation: After all tasks were completed, each participant was asked to fill in the questionnaire for subject evaluation.

5.4 Questions and Methods

Altogether, there are three important questions to be answered by the experiment:

- ”Can elderly use the MIGSEP system to complete the tasks?”
  Besides a general assessment of the task success, a standard measurement method Kappa coefficient ([4]) is also used to detail the evaluation of the effectiveness.
- “Can elderly persons handle the tasks with MIGSEP efficiently?”
  This is answered by the automatically logged data of every single interaction.
- “Do elderly find it comfortable to interact with MIGSEP?”

This is answered by the data of the evaluation questionnaires.

6 Results and Discussion

6.1 Effectiveness of MIGSEP

Regarding the effectiveness of the MIGSEP system, 326 out of 341 tasks (10.5 of 11 for each, 95.6%) were correctly performed by all the participants, where 10.6 (96.6) and 10.4 (94.5%) tasks were completed by each participant using touch or speech input modalities respectively. This suggests a generally high effectiveness of the interaction with the MIGSEP system. However, in order to assess the effectiveness at a more detailed level, the standard statistical method Kappa coefficient was used.

In order to apply the kappa method, we needed to define the attribute value matrix (AVM), which contains all information that has to be exchanged between MIGSEP and the participants. E.g. table 1 shows the AVM for the task: ”Drive to a person named Michael Frieling.” for both touch and speech modalities, where the expected values of this task are also presented.

Table 1. An example AVM for the task “drive to a person name Michael Frieling”.

Touch		Speech	
Attribute	Expected value	Attribute	Expected value
Reached Level	L1, L2, L3, L4	FN	Michael
Goal Selection	Michael Frieling	LN	Frieling
Confirm	Yes (to the correct goal)	G	Male
		M	Person

By combining the actual data recorded during the experiment with the expected attribute values in the AVMs, we can construct the confusion matrices for all tasks. Table 2 shows e.g. the confusion matrix for the task ”drive to a person named Michael Frieling” with the speech input modality, where ”M” and ”N” denote whether the actual data match with the expected attribute values in the AVMs, and “SNU” for the system failed-understanding situation. E.g. first name of Michael is wrongly targeted for 4 times and wrongly understood by the system 14 times. Similar construction is done with the AVM of touch input.

Table 2. The confusion matrix fort the task „drive to a person named Michael Frieling“

	FN			LN			G			M			
Data	M	N	SNU	M	N	SNU	M	N	SNU	M	N	SNU	sum
FN	81	4	14										99
LN				82	3	12							97
G							57		4				61
M										3	2	1	6

Given one confusion matrix, the Kappa coefficient can then be calculated with

$$\kappa = \frac{P(A)-P(E)}{1-P(E)}, \text{ (cf. [4])}$$

In our experiment,

$$P(A) = \frac{\sum_{i=1}^{n} M(i,\ M)}{T}$$

is the proportion of times that the actual data agree with the attribute values, including the system failed-understanding situation as not matched, and

$$P(E) = \sum_{i=1}^{n} \left(\frac{M(i)}{T}\right)^2$$

is the proportion of times that the actual data are expected to be agreed by chance, where M(i, M) is the value of the matched cell of row i, M(i) the sum of the cells of row i, and T the sum of all cells.

Therefore, we summarized the results of all the tasks and constructed one confusion matrix for all the data, and got: kappa coefficient $\kappa = 0.88(std.=0.10)$ for touch and $\kappa = 0.74\ (std.=0.13)$ for speech modality. This in general still suggests a successful degree of interaction using touch and speech modality. However, touch input is performing more effectively at the detailed interaction level compared to the speech modality, due to the common problem caused by the automatic speech recognizer (294 SNU, 19.6 averagely for each participants).

6.2 Efficiency of MIGSEP

Regarding the efficiency of MIGSEP, the automatically logged quantitative data are summarized in table 3, with respect to user turns, system turns and the elapsed time for each participant and each task.

Table 3. Quantitative results calculated based on the recorded data concerning efficiency.

	Touch		Speech	
	Mean	Std.	Mean	Std.
User turns	15.5 (7)	4.1	4.3 (3)	1.7
Sys turns	15.4 (5)	3.9	4.3 (3)	1.6
Elapsed time (s)	88.9	40.2	57.6	24.2

From a general interaction perspective, a very good overall performance efficiency is shown by averagely 4.3 user turns and 4.3 system turns per task for each participant using speech modality, because the average basic turn numbers, inferred by the shortest solution for each task to be filled, are 3 user turns and 3 system turns. This indicates that almost every participant (std. < 2) was able to find the shortest way to complete the tasks while tolerating the problem of automatic speech recognizer. However, 15.5 user turns and 15.4 system turns compared with the shortest solution 7 and 5 respectively were less convincing for the touch modality. Given further insight into the individual data, four participants were having much more turns (averagely 21.8 user turns) than the others, since they were slightly lost finding certain targets and started to do unreasonable brute-force searching.

On the other hand, the elapsed time for each task and each participant for both modalities is considered as satisfying: with averagely 88.9 second for theoretically

minimal 12 interaction paces (7+5) with touch input, which is only 7.4 second each, and 57.5 second for minimal 6 interaction paces (3+3), which is only 9.6 second each. However, the standard deviation of 40.2 for touch input is a bit high, due to the interaction context unawareness of the four individuals with 144.8 second averagely consumed time.

6.3 User Satisfaction

Table 4. The assessment of subjective user satisfaction.

	Touch		Speech	
	Mean	Std.	Mean	Std.
System behavior	4.8	0.3	3.6	0.7
Speech output	4.7	0.5	4.6	0.6
Textual output	4.9	0.3	4.5	0.5
Interface presentation	4.8	0.3	4.7	0.4
Task performing	4.5	0.3	4.3	0.5
User-friendliness	4.7	0.4	4.4	0.6
User perspective	4.3	0.9	3.9	0.8
Overall	4.7	0.2	4.3	0.4

Overall, it shows a very good user satisfaction for touch and speech modalities, with averagely 4.7 and 4.3 out of 5. Specifically, the speech and textual outputs are considered appropriately constructed; the interface is intuitive and easy to understand; the process to perform the task is feasible; and the system is considered user-friendly.

However, the scores of system behavior and user perspective for the speech modality were a bit lower than the others. This is mainly due to the problem of the automatic speech recognizer, which could trigger unexpected system responses, and therefore make the future use from the user perspective less attractive. This impression is also reflected to every other aspect for speech modality, with overall lower score compared to touch input.

7 Conclusions and Future Work

This paper summarized our work on multimodal interaction for elderly persons, centering the following two essential aspects:

- The design and implementation of a multimodal interactive system according to a number of elderly-friendly guidelines concerning with the basic design principles of conventional interactive interfaces and ageing centered characteristics;
- The modelling and development of multimodal interaction using a tool-supported, formally tractable and extensible unified dialogue modelling approach.

In order to evaluate the minutely designed and developed multimodal interactive system, an experimental study was conducted with 31 elderly persons and concerned

with the touch and speech input modality respectively. The evaluation showed high effectiveness, sufficient efficiency and a high satisfaction of the participants with our system for both modalities. Due to the problem caused by the automatic speech recognition, touch modality displays a better performance with respect to effectiveness and is preferred by the elderly. But the speech modality helped the participants in a more efficient way. Thus, the combination of both modalities is motivated.

The presented work served as a continuing step towards building an effective, efficient, adaptive and robust multimodal interactive framework extensively for elderly persons. The result of a further study focusing on the touch and speech combined modality is being analyzed. Corpus-based supervised and reinforcement learning techniques will be applied to improve the current dialogue model and gain more flexible interaction, with the expectation of compensating for the insufficient reliability of automatic speech recognizers.

Acknowledgements. We gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG) through the Collaborative Research Center SFB/TR8, the department of Medical Psychology and Medical Sociology, the department of Neurology of the University Medical Center Göttingen and the St. Mauritius Klink Meerbusch.

References

1. Jaimes, A., Sebe N., 2007. Multimodal human-computer interaction: A survey. In *Computational Vision and Image Understanding. Elsevier Science Inc.,* New York, USA, pp. 116-134.
2. Oviatt, S. T., 1999. Ten myths of multimodal interaction. In *Communications of the ACM.* ACM New York, USA, Vol. 42, No. 11, pp. 74-81.
3. Holzinger, A., Mukasa, K.S., Nischelwitzer, A.K., 2008. Introduction to the special thematic session: Human-computer interaction and usability for elderly. In *Proceedings of the 11th International Conference on Computers Help People with Special Needs.* Springer Verlag, Berlin, Germany, pp. 18-21.
4. Walker, M.A., Litman, D.J., Kamm, C.A., Kamm, A.A., Abella, A., 1997. Paradise: a framework for evaluating spoken dialogue agents. In *Proceedings of the eighth conference on European chapter of Association for computational Linguistics,* NJ, USA, pp. 271-280.
5. Morris, J.M., 1994. User interface design for older adults. In *Interacting with Computers.* Vol. 6, 4, pp. 373-393.
6. Jian, C., Scharfmeister, F., Rachuy, C., Sasse, N., Shi, H., Schmidt, H., Steinbüchel-Rheinwll, N. v., 2011. Towards Effective, Efficient and Elderly-friendly Multimodal Interaction. In *PETRA 2011: Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments. ACM,* New York, USA.
7. Fozard, J.L., 1990. Vision and hearing in aging. In *J. Birren, R. Sloane and G.D. Cohen (eds), Handbook of Metal Health and Aging.* Academic Press, Volume 3, pp. 18-21.
8. Mackay, D., Abrams, L., 1996. Language, memory and aging. In *J.E. Birren and K.W.Schaie (eds), Handbook of the psychology of Aging*. Academic Press, Volume 4, pp. 251-265.

9. Moeller, S., Goedde, F., Wolters, M., 2008. Corpus analysis of spoken smart-home interactions with older users. In N. Calzolari, K.Choukri, B. Maegaard, J. Mariani, J. Odjik, S. Piperidis, and D. Tapias, (eds.), Proceedings of the Sixth International Conference on Language Resources Association. ELRA.
10. Kline, D.W., Scialfa, C.T., 1996. Sensory and Perceptual Functioning: basic research and human factors implications. In *A.D. Fisk and W.A. Rogers. (eds.), Handbook of Human Factors and the Older Adult,* Academic Press.
11. Schieber, F., 1992. Aging and the senses. In *J.E. Birren, R.B. Sloane, and G.D. Cohen, (eds.) Handbook of Mental Health and Aging,* Academic Press, Volume 2.
12. Walkder, N., Philbin, D.A., Fisk, A.D., 1997. Age-related differences in movement control: adjust submovement structure to optimize performance. In *Journal of Gerontology: Psychological Sciences 52B,* pp. 40-52.
13. Charness, N., Bosman, E., 1990. Human Factors and Design. In *J.E. Birren and K.W. Schaie, (eds.), Handbook of the Psychology of Aging.* Academic Press, Volume 3, pp. 446-463.
14. Kotary, L., Hoyer, W.J., 1995. Age and the ability to inhibit distractor information in visual selective attention. In *Experimental Aging Research.* Volume 21, Issue 2.
15. McDowd, J.M., Craik, F. 1988. Effects of aging and task difficulty on divided attention performance. In *Journal of Experimental Psychology: Human Perception and Performance 14*. pp. 267-280.
16. Hoyer, W.J., Rybash, J.M., 1992. Age and visual field differences in computing visual spatial relations. In *Psychology and Aging 7.* pp. 339-342.
17. Salthouse, T.A., 1994. The aging of working memory. In *Neuropsychology 8,* pp. 535-543.
18. Craik, F., Jennings, J., 1992. Human memory. In *F. Craik and T.A. Salthouse, (eds.), The Handbook of Aging and Cognition.* Erlbaum, pp. 51-110.
19. Shaie, K.W., 1996. Intellectual development in adulthood. In *J.E. Birren and K.W. Shaie, (eds.), Handbook of the psychology of aging.* Academic Press, Volume 4.
20. Sitter, S., Stein, A., 1992. Modelling the illocutionary aspects of information-seeking dialogues. In *Journal of Information Processing and Management.* Elsevier, Volume 28, issue 2, pp. 165-180.
21. Traum, D., Larsson, S., 2003. The information state approach to dialogue management. In *J.v. Kuppevelt and R. Smith (eds.), Current and New Directions in Discourse and Dialogue.* Kluwer, pp. 325-354.
22. Ross, J.R., Bateman, J., Shi, H., 2005. Using Generalized Dialogue Models to Constrain Information State Based Dialogue Systems. In *Symposium on Dialogue Modelling and Generation.*
23. Alston, P.W., 2000, *Illocutionary acts and sentence meaning.* Cornell University Press.
24. Roscoe, A.W., 1997. *The Theory and Practice of Concurrency,* Prentice Hall.
25. Hall, A., Chapman, R., 2002. Correctness by construction: Developing a commercial secure system. In *IEEE Software.* Vol. 19, 1, pp. 18-25.
26. Shi, H., Bateman, J., 2005. Developing human-robot dialogue management formally. In Proceedings of Symposium on Dialogue Modelling and Generation. Amsterdam, Netherlands.
27. Broadfoot, P., Roscoe, B., 2000. Tutorial on FDR and Its Applications. In *K. Havelund, J. Penix and W. Visser (eds.), SPIN model checking and software verification.* Springer-Verlag, London, UK, Volume 1885, pp. 322.
28. Folstein, M., Folstein, S., Mchugh, P., 1975. “mini-mental state”, a practical method for grading the cognitive state of patients for clinician. In *Journal of Psychiatric Research.* Volume 12, 3, pp. 189-198.

Better Choice? Combining Speech And Touch In Multimodal Interaction For Elderly Persons

Cui Jian[1], Hui Shi[1], Nadine Sasse[2], Carsten Rachuy[1], Frank Schafmeister[2], Holger Schmidt[3] and Nicole von Steinbüchel[2]

[1]SFB/TR8 Spatial Cognition, University of Bremen, Enrique-Schmidt-Straße 5, Bremen, Germany
[2]Medical Psychology and Medical Sociology, University Medical Center Göttingen, Waldweg 37, Göttingen, Germany
[3]Neurology, University Medical Center Göttingen, Robert-Koch-Str. 40, Göttingen, Germany
{ken, shi}@informatik.uni-bremen.de, n.sasse@med.uni-goettingen.de, rachuy@informatik.uni-bremen.de, {frank-schafmeister, h.schmidt, nvsteinbuechel}@med.uni-goettingen.de

Keywords: ICT, Ageing and Disability; Elderly-centered Design; Multimodal Interaction; Elderly-friendly Interface; Formal Methods; System Evaluation.

Abstract: This paper presents our work on developing, implementing and evaluating a multimodal interactive guidance system that features spoken language and touch-screen input for elderly persons. The development foundation of the system comprises two systematically designed and empirically improved aspects: a set of development guidelines for elderly-friendly multimodal interaction according to common ageing-related decline of important human abilities, and a hybrid dialogue modelling approach with a formal method triggering and agent-based management for the elderly-centered multimodal interaction. To evaluate the minutely developed and implemented system, an experimental study was conducted with thirty-three elderly persons and empirical data were analyzed by applying an adapted version of a general evaluation framework, which provided overall positive analysis results and validated our effort to develop an effective, efficient and elderly friendly multimodal interaction.

1 INTRODUCTION

As the demographic development shows, the amount of elderly people is constantly growing in modern societies (Lutz, Sanderson and Scherbov, 2008). These persons often suffer from age-related decline or impairment of sensory, perceptual, physical and cognitive abilities. This poses particular challenges to the application of technical systems nowadays, which are getting more and more commonly implemented in the daily routines of elderly persons.

Meantime, attention is increasingly focused on the technical systems with multimodal interfaces, which provide the users with multiple modes of interaction with a system; therefore they improve the quality of human-system communication concerning effectiveness, efficiency and user-friendliness (cf (Jaimes and Sebe, 2007)).

Thus, in order to maximize the usability and user experience of technical systems for elderly persons, research on multimodal interaction for this specific user group is increasingly gaining more interest during the last decade. Various emerging technologies have been considered, such as advanced speech enabled interface (Krajewski, Wieland and Batliner, 2008), brain-signal interface (Mandel, et al., 2009), visual input via digital camera (Goetze, et al., 2012); also, a large contribution is being made to "Ambient Assistive Living", the concept for developing age-adjusted and care-friendly living environments (cf. (Rodríguez, García-Vázquez and Andrade, 2011)).

This paper presents a multimodal interactive system that can provide elderly persons with both spoken language and touch-screen input modalities. It has been particularly developed and implemented for the elderly focussing on two important aspects: 1) a set of development guidelines for multimodal interactive systems with respect to the basic design principles of conventional interactive systems and the most common age-related characteristics; and 2) a hybrid dialogue modelling and management approach that combines the advanced finite state

based generalized dialogue model and the classic agent based dialogue theory; it supports a flexible and context-sensitive, yet tractable and controllable multimodal interaction with a formal language based development framework. The resulting system has been continuously improved by a series of evaluative studies of our previous work concerning the development foundation and different modalities (cf. (Jian, et al, 2011), (Jian, et al, 2012)). In order to perform a further evaluation of spoken language and touch-screen combining input modalities, as well as the assessment of the complete multimodal interactive system concerning its effectiveness of task performance, efficiency of interface interaction and user acceptance by elderly persons, a supplementary experimental study was conducted with 33 elderly. The data were analysed by applying an adapted version of the general evaluation framework PARADISE (Walker, et al., 1997). The Results are briefly described in this paper.

The rest of the paper is organized as follows: section 2 briefly introduces the design guidelines for multimodal interactive systems for elderly persons and the hybrid approach for multimodal interaction management; section 3 presents the multimodal interactive guidance system and section 4 describes the experimental study on evaluating the modality combining spoken language and touch-screen; the results are analysed and discussed in section 5. Finally, section 6 concludes the reported work and outlines the direction of our future activities.

2 THE DEVELOPMENT FOUNDATION

The theoretical and technical foundation of our work comprises two aspects:

- A set of design guidelines for an elderly-centered multimodal interactive system;
- A formal method and agent based hybrid modelling approach for dialogue management.

They are both systematically designed with respect to their suitability for the application, and continuously improved by our previous empirical studies (cf. (Jian, et al, 2012)).

2.1 Design Guidelines of Multimodal Interactive Systems for Elderly Persons

Physical and cognitive decline is almost universal in the elderly. According to (Birdi, Pennington and Zapf, 1997), these age-related characteristics should be considered while developing interactive systems for the elderly. Therefore, based on the common design principles for conventional interactive systems and the ageing-related empirical findings, we defined a set of design guidelines for multimodal interaction with respect to the decline of important human perceptual and cognitive functions. These guidelines have been implemented into our multimodal interactive guidance system, evaluated by our previous empirical studies, and then improved on the basis of their results.

The final set of improved design guidelines were summarized in (Jian, et al, 2012). For reasons of brevity we report empirical findings regarding the decline of the seven most common human abilities, accordingly followed by the most important elements implemented and improved during our system development:

Visual Perception worsens for most people with age (Fozard, 1990). Physically the size of the visual field is decreasing and the peripheral vision can be lost. It is more difficult to focus on objects up close and to see fine details, including rich colours and complex shapes that make images hard or even impossible to identify. Rapidly moving objects are either causing too much distraction, and/or become less noticeable. This decline concerns most with the graphical user interface. Based on the suggested guidelines, only simple and clear layout was constructed without overlapping items; 12-14 sized sans-serif fonts were chosen for all displayed texts. Simple and high contrast colours without fancy visual effects were used and placed aside; regularly shaped rectangles and circles were selected for comfortable perception and easy identification.

Speech Ability declines while ageing in the way of being less efficient for pronouncing complex words or longer sentences, probably due to reduced motor control of tongue and lips (Mackay and Abrams, 1996). (Moeller, Goedde and Wolters, 2008) confirmed that, elderly-centered adaptation of speech-enabled interactive components can improve the interaction quality to a satisfactory level. Therefore, the vocabulary and grammar for our speech recognizer and analyser were constructed with preferably short and easy wording in daily life communication; dialogue strategies were also adjusted to many elderly-specific situations.

Auditory Perception declines to 75% between the age of 75 and 79 year olds (cf. Kline and Scialfa, 1996). High pitched sounds are hard to perceive; complex sentences are difficult to follow (Schieber, 1992). Therefore, text and acoustic output were both

provided as system responses. Style, vocabulary and structure of the sentences were intensively revised regarding brevity. A low-pitched yet vigorous male voice was used for the speech synthesis.

Motor Abilities decline generally due to loss of physical activities while ageing. Complex fine motor activities are more difficult to perform, e.g. to grab small or irregular targets (cf. Charness and Bosman, 1990); conventional input devices such as a computer mouse are less preferred by elderly persons as good hand-eye coordination is required (Walkder, Philbin and Fisk, 1997). Taking these findings into account, a touch screen was chosen as the haptic interface; Regularly shaped, sufficiently sized and well separated interface elements were constructed; pressing instead of clicking or dragging was decided to be the only action in order to avoid otherwise frequently occurring errors.

Attention and Concentration drop while ageing. Elderly persons either become more easily distracted by details and noise, or find other things harder to notice when concentrating on one thing (Kotary and Hoyer, 1995); they show great difficulty with situations where divided attention is needed (McDowd and Craik, 1988). Thus, fancy irrelevant images or decorations were removed. Unified font, colours, sizes of interface elements were used throughout the interaction. Simple animations for notifying changes were constructed, giving sufficiently clear feedback to the user.

Memory Functions decline differently. Short term memory holds fewer items with age and working memory becomes less efficient (Salthouse, 1994). Semantic information is normally preserved in long term memory (Craik and Jennings, 1992). Guided by these facts, the quantity of displayed items was restricted to no more than three, regarding the average capacity of short term memory of elderly persons; sequentially presented items were intensively revised to assist orientation during interaction. Context sensitive cues were presented with selected colours: green for items concerning persons, yellow for items concerning rooms, etc.

Intellectual Reasoning Ability does not decline much during the normal ageing process. (Hawthorn, 2000) believed that crystallized intelligence can assist elderly persons to perform better in a stable well-known interface environment. Therefore, consistent layout, colours and interaction styles were used throughout the interaction. Changes on the interface can only happen on data level.

2.2 The Hybrid Approach for Interaction Management

The hybrid dialogue modelling approach combines the finite-state-based generalized dialogue models with the classic agent-based dialogue management theories. This section

- introduces the basic theory of this approach,
- presents the adapted instance model for multimodal interaction in elderly persons by applying the hybrid approach;
- describes a formal language based development toolkit, which is then used to support the implementation of the instance model and its integration into our multimodal interactive guidance system for achieving an effective, flexible, yet formally controllable multimodal interaction management.

2.2.1 The Theory

The development of the hybrid dialogue modelling approach benefited from existing researches on these two important interaction management theories:

The **generalized dialogue models**, which are constructed with recursive transition networks (RTN) at the illocutionary level. These networks can abstract dialogue models by describing discourse patterns as illocutionary acts, without reference to any direct surface indicators (cf. (Alston, 2000));

The **classic agent-based management** method: information state update based management theories (cf. (Traum and Larsson, 2003)), which focus on the modelling of discourse context as the attitudinal state of an intelligent agent. This method shows a powerful way to handle dynamic information for a context sensitive dialogue management.

However, these two well-accepted methods have their own limitations. On one hand, the generalized dialogue models are based on finite state transition models, which are criticized for their inflexibility of dealing with dynamic information exchange; on the other hand, the information state update models are usually very difficult to manage and extend.

Therefore, we designed a hybrid dialogue modelling approach by extending the generalized dialogue model with conditions and information state update rules added into finite-state transitions.

2.2.2 The Interaction Model

In order to manage multimodal interaction for elderly persons, an adapted hybrid dialogue model was constructed and evaluated by our previous work (cf. (Shi, Jian and Rachuy, 2011)). The accordingly improved version consists of four hybrid dialogue schemas: the initiating schema, the user's action schema, the system's response schema and the

user's response schema, regarding the four general transitions during interaction.

Each interaction is initiated with the schema *Dialogue(S, U)* (cf. Figure 1), by the initialization of the system's start state and a greeting request. In Dialogue(S, U) the system initiates a dialogue with a request move (i.e. *S.request*), which cause the initialization of the dialogue context using the update rule INIT.

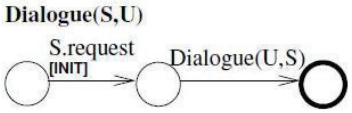


Figure 1: the initiating model.

The dialogue continues with the user's instruction, request for a certain information or restart action, leading to the system's further response or dialogue restart, respectively, while updating the information state with the attached update rules (cf. *Dialogue(U, S)* in figure 2).

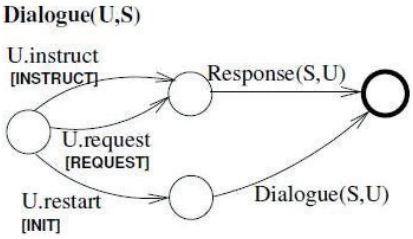


Figure 2: the user's actions model.

After receiving user input, the system tries to generate an appropriate response according to its current knowledge base and information state (cf. *Response(S, U)* in figure 3). This can be informing the user with requested data, rejecting an unacceptable request with or without certain reasons, providing choices for multiple options, or asking for further confirmation of taking a critical action, each of which triggers transitions to other hybrid models.

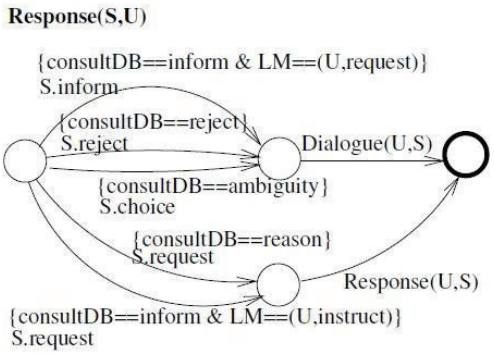


Figure 3: the system's response model.

Finally, the user can accept or reject the system's response, or even ignore it by simply providing new instructions or requests, triggering further state transitions as well as information state updates (cf. *Response(U, S)* in Figure 4).

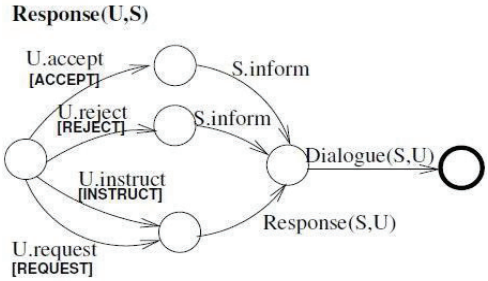


Figure 4: the user's response model.

Besides the improvement performed with respect to the specific interaction data of the elderly subjects in our previous studies, the decline of physical and cognitive abilities of elderly persons, especially memory function, concentration and fluid reasoning ability should be considered as well. Therefore, for the improvement of the current hybrid dialogue model we also included the following features to assist the elderly during the multimodal interaction:

- Relevant dialogue history information, such as the latest utterance, was added into the current information state and provided in case of speech recognition problems.
- Context sensitive information, which is kept in the current information state, is designed to be either directly presented after each interaction pace, or included within dialogue utterance, in order to ease the common problems caused by the declining memory function.
- Additional context information is provided with specific information state update rules in extreme cases, e.g. if the automatic speech recognition problems become too interfering, messages containing possibly recognized context will be presented.
- Instead of keeping rich transition alternatives at the illocutionary level, the hybrid model was kept as compact and intuitive as possible.

2.2.3 A Development Framework to Support the Hybrid Dialogue Modelling Approach

The structure of a hybrid dialogue model is in fact a typical finite state transition model. This feature enables any hybrid dialogue model to be formally specified as a set of machine readable codes, e.g.

using mathematically well-founded formal language, Communicating Sequential Processes (CSP) (cf. (Roscoe, 1997)) in the formal methods and computer science community. Furthermore, the CSP program is also supported by well-established model checkers, which provides the rich possibilities of validating the concurrent aspects and increasing the tractability of the specified model (cf. (Hall, 2002)).

Thus, in order to support the development of hybrid dialogue models using the formal language CSP and its integration into a practical multimodal interactive system, we designed FormDia, the Formal Dialogue Development Toolkit (cf. Figure 5).

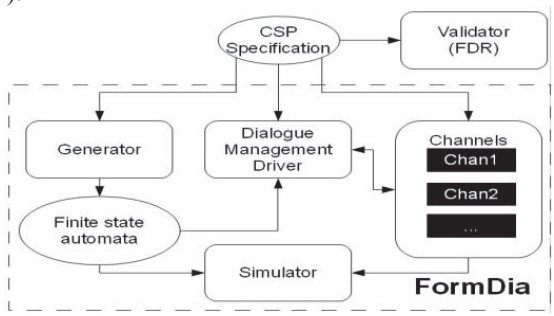


Figure 5: the Structure of the FormDia Toolkit.

Theoretical and technical details about FormDia can be found in (Shi and Bateman, 2005). In general, the FormDia Toolkit supports the implementation and integration of a hybrid dialogue model into an interactive system with the four components:

Validator: after a hybrid dialogue model is specified with CSP, it can be validated by an external model checker: the Failures-Divergence Refinement tool or FDR (Broadfoot and Roscoe, 2000), for validating and verifying concurrency of state automata.

Generator: with the validated CSP specification, machine readable finite state automata can be generated by the Generator.

Simulator: with the generated finite state automata and the communication channels, dialogues scenarios can be simulated via a graphical interface, which visualizes dialogue states as a directed graph and provides a set of utilities for primary testing.

Dialogue Management Driver: finally the dialogue model is integrated into an interactive system via the dialogue management driver.

Therefore, FormDia enables an intuitive design of hybrid dialogue models with formal language, automatic validation of the related functional properties, easy simulation and verification of specified interaction situations, and a straightforward integration into a practical interactive system.

3 SYSTEM DESCRIPTION

Based on the development foundation introduced in the previous section, we developed a general Multimodal Interactive Guidance System for Elderly Persons (MIGSEP).

3.1 System Introduction

MIGSEP runs on a portable touch-screen tablet PC and will serve as the interactive media, which is intended to be used by an elderly or handicapped person seated in an autonomous electronic wheelchair that can automatically carry its users to desired locations within complex environments. The user should interact with MIGSEP with spoken-language and touch-screen combining input modality to find the desired target.

3.2 System Architecture

The architecture of MIGSEP is illustrated in figure 6. The *Generalized Dialogue Manager* was developed using the introduced adapted hybrid dialogue model and the FormDia toolkit. It functions as the central processing unit of the entire system and supports a formally controllable and extensible, meantime flexible and context-sensitive multimodal interaction management. An *Input Manager* receives and interprets all incoming messages from the *GUI Action Recognizer* for GUI input events, the *Speech Recognizer* for natural language understanding and the *Sensing Manager* for other possible sensor data. An *Output Manager* on the other hand, handles all outgoing commands and distributes them to the *View Presenter* for presenting visual feedbacks, the *Speech Synthesizer* to generate natural language responses and the *Action Actuator* to perform necessary motor actions, such as sending a driving request to the autonomous electronic wheelchair. The *Knowledge Manager*, constantly connected with the *Generalized Dialogue Manager,* uses a *Database* to keep the static data of certain environments and the *Context* to process the dynamic information exchanged with the users during the interaction.

All components of MIGSEP are closely connected via XML-based communication channels and each component can be treated as an open black box which can be accordingly modified or extended

for concrete domain specific use, without affecting other components in the MIGSEP architecture. It provides a general open platform for both theoretical researches and empirical studies on single- or multimodal interaction that can relate to different application domains and scenarios.

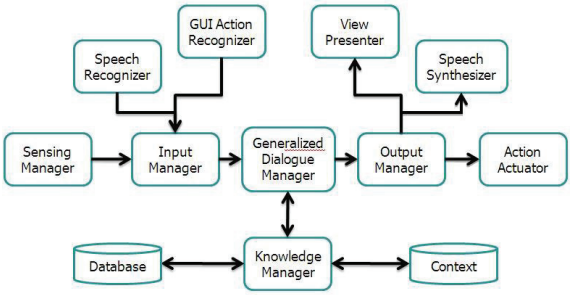


Figure 6: The architecture of MIGSEP.

3.3 Interaction with the System

The current instance of MIGSEP was implemented as a guidance system used by elderly persons for the application domain of hospital environments. Figure 7 shows a user interacting with MIGSEP.

Figure 7: a user is interacting with MIGSEP.

This MIGSEP system consists of a button device for triggering a "press to talk" signal, a green lamp to signalize the "being pressed and ready to talk" state, and the tablet PC, on which the MIGSEP system is running and the interface is displayed. The MIGSEP interface simply consists of two areas:

Function-area contains the function button "start" on the top left for going to the start state, the function button "toilet" below it regards the basic needs of elderly persons, and the text area for displaying the system responses in the middle;

Choice-area displays information entities as single cards that can be selected, with a scrollbar indicating the position of the current displayed cards and a context sensitive coloured bar showing the current concerned context if necessary.

Figure 8 shows a sample of spoken language and touch-screen combined interaction dialogue between MIGSEP and a user who would like go to the cardiology department, to a doctor named Wolf.

S: <shows 3 cards, green for persons, yellow for rooms and blue for departments>
hello, where do you want to go?
U: I want to go to Wolf in the cardiology department.
S: <shows a blue bar texted with cardiology on the top, and 2 green cards, with photos, titles and names in the middle>
I found two persons working in the cardiology department with the name Wolf, to which do you want to go?
U: <presses the second card>
S: <enlarges the second green card, shows yes/no button>
Do you want to go to Mrs. Dr. Diana Wolf?
U: <press the yes button>
S: <resizes the second green card and shows the 3 cards in the start menu again>
OK, I have saved this goal. Where else do you want to go?

Figure 8: a sample interaction with MIGSEP.

4 EXPERIMENTAL STUDY

To evaluate how well the MIGSEP system can assist elderly persons by using a modality combining spoken language and touch-screen, an experimental study was conducted with the department of Medical Psychology and Medical Sociology in Göttingen.

4.1 Participants

33 elderly persons (m/f: 19/14, mean age of 70.7, standard deviation 3.1), all German native speakers, participated in the study. They all had to pass the mini-mental state examination (MMSE), a screening test to assess the cognitive mental status (cf. (Folstein, Folstein and Mchugh, 1975)). A test score between 28 and 30 indicates slightest decline versus normal cognitive functioning. Our participants showed an average score of 28.9 (std.=.83).

4.2 Stimuli and Apparatus

Visual stimuli were presented via a green lamp and a graphical user interface on the screen of a portable tablet PC; audio stimuli were also generated by the MIGSEP system and played via two loudspeakers at a well-perceivable volume. All tasks were given as keywords on the pages of a calendar-like system.

There were two types of input possibilities, which could be freely chosen: the spoken language, activated if the button was being pressed and the green lamp was on; and the touch-screen action, directly performable on the touch-screen display.

The same data set contains virtual yet sufficient information about personnel, rooms and departments in a common hospital, was used in the experiment.

During the experiment each participant was accompanied by the same investigator, who

introduced the system and gave well-defined instructions at the beginning, and provided help if necessary during the trail (which was very rare).

An automatic internal logger of the MIGSEP system was used to collect the real-time system internal data, while the windows standard audio recorder program kept track of the whole dialogic interaction process.

A questionnaire, focusing on the user satisfaction with MIGSEP with respect to the spoken language and touch-screen combining input modality, was especially designed for this study. It contains 6 questions concerning the quality of the combined modality compared to a single modality, the feasibility, the advantages, the usability, the appropriateness and the preference. This questionnaire was answered by each participant via a five point Likert scale.

4.3 Procedure

Each participant had to undergo four phases:

Introduction: a brief introduction was given to the participants, so that they could get the basic idea and an overview of the experiment.

Learning and Pre-tests: the participants were instructed how to interact with MIGSEP using the spoken natural language and the touch-screen input. In order to minimize the learning or bias effect with respect to the use of one modality, we introduced a cross-over procedure, 16 participants out of 33 had to first use the touch-screen input and then the spoken language, the other 17 used spoken language first and then the touch-screen input. All of the participants had to perform 11 tasks concerning their navigation procedures in a hospital in order to reach a certain aim. Each modality and each task contained incomplete yet sufficient information about a destination the participant should select. For example they had to drive to "room 2603", to "Sonja Friedrich", or to "room 1206 or room 2206 with the name OCT-Diagnostics". Tasks were fulfilled or ended, if the goal was selected or the participant gave up trying after six minutes.

Testing: After performing 22 tasks with both modalities, each participant was asked to freely choose between spoken language and touch-screen input modality to perform again 11 tasks; they contained similar information as in pre-tests (varied only on data level) and were performed under the same conditions.

Evaluation: After all tasks were run through, each participant was asked to fill in the questionnaire for evaluation.

5 RESULTS AND ANALYSIS

According to the Paradise framework (Walker, et al., 1997), the performance of an interactive system can be measured via the effectiveness, efficiency of the system and the user satisfaction. Therefore, these three aspects were analysed.

5.1 Effectiveness of the System

To find out how effective the elderly were assisted by the MIGSEP system with the combining modality, statistical method "Kappa coefficient" is used. However, in the classic Paradise framework the Kappa method was originally used to evaluate a spoken dialogue system.

Therefore, in order to be able to calculate the Kappa coefficient with respect to the multimodal interaction with the MIGSEP system, we first had to develop an adaptation of the original attribute value matrix, which still contained all information that was exchanged during the multimodal interaction between MIGSEP and participants. For this reason, we introduce the concept of an Attribute Value Tree (AVT) (cf. the example in Figure 9).

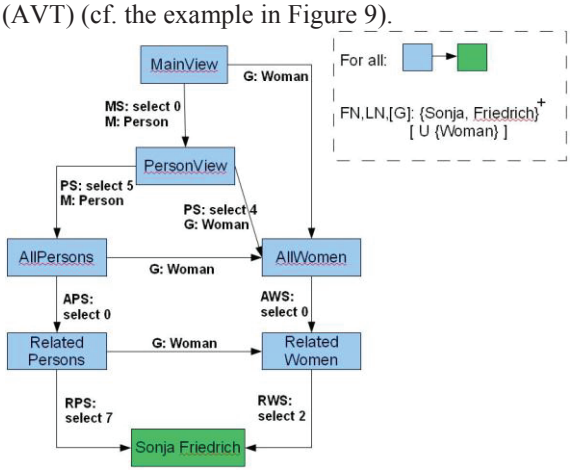


Figure 9: an Attribute Value Tree.

An AVT is defined as a finite state transition diagram, which contains all the expected correct way, either touch-screen input or spoken language command, as the state transitions from the start state to the target state. As the AVT for the task "go to a person named Sonja Friedrich" illustrated in Figure 9, any correct interaction should go from the state mainView, then e.g. to PersonView by selecting the first card (MS: select 0), or performing the spoken language command "I want to go to a person" (M: Person), or go to AllWomen state by simply saying "I want to go to a woman" from the MainView, etc.

An AVT contains the expected data set of a task, and therefore functions similarly as the original

attribute value matrix, yet with the possibility of recording multimodal interaction exchange.

Thus, 11 AVTs were created for the 11 tasks respectively and by combining the actual data recorded during the experiment with the expected attribute values in the AVTs, we can construct the confusion matrices for all tasks. E.g., table 1 shows the confusion matrix for the task ”drive to a person named Sonja Friedrich”, where ”M” and ”N” denote whether the actual data match with the expected attribute values in the AVTs. E.g. there were 25 correctly selected actions in the PersonSelect (PS) state; and the spoken language command regarding the first name (FN) was misrecognized by the system for 6 times. Note that, because of the width of the text, not all attributes of this confusion matrix can be shown in this example.

Table 1: The confusion matrix for the task “drive to a person named Sonja Friedrich”.

	PS		MS		...		FN		
Data	M	N	M	N	M	N	M	N	sum
PS	25	0							25
MS			14	0					14
...					...	...			...
FN							62	6	68

The data for all confusion matrices were merged and a total confusion matrix for all the data of the 11 performed tasks was created.

Given the total confusion matrix, the Kappa coefficient was calculated with

$\kappa = \frac{P(A)-P(E)}{1-P(E)}$, (Walker, et al., 1997)

In our experiment, $P(A) = \frac{\sum_{i=1}^{n} M(i,\ M)}{T}$ is the proportion of times that the actual data agree with the attribute values, and $P(E) = \sum_{i=1}^{n}(\frac{M(i)}{T})^2$ is the proportion of times that the actual data are expected to be agreed on by chance, where M(i, M) is the value of the matched cell of row i, M(i) the sum of the cells of row i, and T the sum of all cells.

Thus, we could calculate the Kappa Coefficient of the total confusion matrix $\kappa=0.91$, suggesting a highly successful degree of interaction between the MIGSEP and the participants using the spoken language and touch-screen combining modality.

5.2 Efficiency of the System

In order to find out how efficiently the participants were assisted using the combining modality, quantitative data of every single interaction during the testing were automatically logged. Results are summarized in Table 2, with respect to four important aspects for efficiency analysis.

Table 2: Efficiency of the system for each participant and each task.

	Average	Standard deviation
User turns	7.4	3.6
Sys turns	7.4	3.6
ASR error times	0.3	0.4
Elapsed time (s)	48.7	20.0

The average 7.4 user turns and 7.4 system turns per participant per task have shown a very satisfying efficiency of the system, because the average basic turn numbers, which can be inferred with the theoretically shortest solution, are 2.9 user turns and 2.9 system turns for the only spoken language input, and 5.6 user turns and 5.3 system turns for the touch-screen input. The standard deviation 3.6 even indicates that, some of the participants are solving tasks using the approximately shortest solutions.

However, as observed the average turn numbers are a bit higher than the average number for the shortest solution, with a further insight into the detailed data, two reasons can be concluded:

- 4 participants were using only touch-screen input to interact with the system, which significantly increased the total turn numbers.
- By combining spoken language and touch-screen inputs, many participants first used the touch-screen to sort out the rough direction for each task and then used spoken language instructions to find the target, which however inevitably increased the turn numbers, yet clearly indicates their intention of avoiding the possible problems caused by the automatic speech recognition. This is also reflected by the very good average ASR error rate: 0.3, no ASR error occurred during the interaction.

Meanwhile, the average elapsed time for each task and participant (48.7 seconds) is considered as very short as well, because even with the shortest solution using spoken language commands, merely 48.7 seconds were used for 5.8 user interaction paces (2.9 user turns + 2.9 system turns), which is averagely maximum 8.4 seconds for each turn, this even includes long sentences uttered by the system for over 10 seconds. Although the standard deviation 20.0 is a bit high, this is caused by the same participants, especially the one who was using only touch-screen and doing brute-force searching, and used averagely 135.8 seconds for each task.

5.3 User Satisfaction of the System

Regarding the user satisfaction of the system, we analysed the subjective data coming from the evaluation questionnaire concerning the interaction with the system with the combining modality. The results are summarized in Table 3, underlining very good user experiences with the combining modality.

Table 3: Data concerning subjective user satisfaction.

	Mean	Standard deviation
Better than single modality?	4.4	1.1
Easier solving tasks?	4.0	1.3
Showing advantages?	4.5	1.0
Usable to use combi-modality?	4.1	1.5
Prefer to use combi-modality?	4.4	1.3
Not confusing?	4.5	0.9
Overall	4.3	1.0

However, the scores of easier solving tasks and the usability of the combining modality were a bit lower than the others and the corresponding standard deviations were also higher. It is again mainly due to the extreme cases, where the participants only used touch-screen input and had made unpleasant impression of using only touch-screen, and therefore gave comparably lower score in the questionnaire.

6 CONCLUSION AND FUTURE RESEARCH

In this paper we reported our work on multimodal interaction for elderly persons by focusing on the following two important aspects:

- The summary of our systematically designed and empirically improved foundation for developing and implementing the elderly-centered multimodal interaction;
- The evaluation of the spoken language and touch-screen combined input modality of a multimodal interactive guidance system for the elderly by applying an adapted well-established evaluative framework.

Results of the evaluation showed a very high degree of effectiveness, efficiency and user satisfaction of our system, specifically by using the spoken language and touch-screen combining input modality. This confirmed our theoretical and technical foundation, approaches and frameworks on developing effective, efficient and elderly-friendly multimodal interactive systems.

The reported work continued the pursuit of our goal towards building effective, efficient, adaptive and robust multimodal interactive systems and frameworks for elderly in ambient assistive living environments. Further studies are needed to investigate the reported extreme cases. Corpus-based supervised and reinforcement learning techniques will be applied to support and improve the formal language driven and agent-based hybrid modelling and management approach. More relevant research and experiments on assisting elderly in navigating through complex buildings are also being conducted.

ACKNOWLEDGEMENTS

We greatly acknowledge the support of the German Research Foundation through the Collaborative Research Center SFB/TR 8 Spatial Cognition, the department of Medical Psychology and Medical Sociology and the department of Neurology of the University Medical Center Göttingen.

REFERENCES

Alston, P.W., 2000, *Illocutionary acts and sentence meaning.* Cornell University Press.

Birdi, K., Pennington, J., Zapf, D., 1997. Aging and errors in computer based work: an observational field study. In *Journal of Occupational and Organizational Psychology.* pp. 35-74.

Broadfoot, P., Roscoe, B., 2000. Tutorial on FDR and Its Applications. In *K. Havelund, J. Penix and W. Visser (eds.), SPIN model checking and software verification.* Springer-Verlag, London, UK, Volume 1885, pp. 322.

Charness, N., Bosman, E., 1990. Human Factors and Design. In *J.E. Birren and K.W. Schaie, (eds.), Handbook of the Psychology of Aging.* Academic Press, Volume 3, pp. 446-463.

Craik, F., Jennings, J., 1992. Human memory. In *F. Craik and T.A. Salthouse, (eds.), The Handbook of Aging and Cognition.* Erlbaum, pp. 51-110.

Folstein, M., Folstein, S., Mchugh, P., 1975. “mini-mental state”, a practical method for grading the cognitive state of patients for clinician. In *Journal of Psychiatric Research.* Volume 12, 3, pp. 189-198.

Fozard, J.L., 1990. Vision and hearing in aging. In *J. Birren, R. Sloane and G.D. Cohen (eds), Handbook of Metal Health and Aging.* Academic Press, Volume 3, pp. 18-21.

Goetze, S., Fischer, S., Moritz, N., Appell, J.E., Wallhoff, F., 2012. Multimodal Human-Machine Interaction for Service Robots in Home-Care Environments. In *Proceedings of the 1st International Conference on*

Speech and Multimodal Interaction in Assistive Environments. The Association for Computer Linguistics, pp. 1-7.

Hall, A., Chapman, R., 2002. Correctness by construction: Developing a commercial secure system. In *IEEE Software.* Vol. 19, 1, pp. 18-25.

Hawthorn, D., 2000. Possible implications of ageing for interface designer. In *Interacting with Computers.* pp. 507-528.

Jaimes, A., Sebe N., 2007. Multimodal human-computer interaction: A survey. In *Computational Vision and Image Understanding. Elsevier Science Inc.,* New York, USA, pp. 116-134.

Jian, C., Scharfmeister, F., Rachuy, C., Sasse, N., Shi, H., Schmidt, H., Steinbüchel-Rheinwll, N. v., 2011. Towards Effective, Efficient and Elderly-friendly Multimodal Interaction. In *PETRA 2011: Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments*. *ACM,* New York, USA.

Jian, C., Scharfmeister, F., Rachuy, C., Sasse, N., Shi, H., Schmidt, H., Steinbüchel-Rheinwll, N. v., 2012. Evaluating a Spoken Language Interface of a Multimodal Interactive Guidance System for Elderly Persons. In *HealthInf 2012: Proceedings of the International Conference on Health Informatics. SciTepress,* Vilamoura, Algarve, Portugal.

Kline, D.W., Scialfa, C.T., 1996. Sensory and Perceptual Functioning: basic research and human factors implications. In *A.D. Fisk and W.A. Rogers. (eds.), Handbook of Human Factors and the Older Adult,* Academic Press.

Kotary, L., Hoyer, W.J., 1995. Age and the ability to inhibit distractor information in visual selective attention. In *Experimental Aging Research.* Volume 21, Issue 2.

Krajewski, J., Wieland, R., Batliner, A., 2008. An acoustic framework for detecting fatique in speech based human computer interaction. In *Proceedings of the 11th International Conference on Computers Help People with Special Needs*. Springer-Verlag Berlin, Heidelberg, pp. 54-61.

Lutz, W., Sanderson, W., Scherbov, S., 2008. The coming acceleration of global population ageing. In *Nature*. pp. 716-719.

Mackay, D., Abrams, L., 1996. Language, memory and aging. In *J.E. Birren and K.W.Schaie (eds), Handbook of the psychology of Aging*. Academic Press, Volume 4, pp. 251-265.

Mandel, C., Lüth, T., Laue, T., Röfer, T., Gräser, A., Krieg-Brückner, B., 2009. Navigating a Smart Wheelchair with a Brain-Computer Interface Interpreting Steady-State Visual Evoked Potentials. In *Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE Xplore, St. Louis, Missouri, United States, pp. 1118-1125.

McDowd, J.M., Craik, F. 1988. Effects of aging and task difficulty on divided attention performance. In *Journal of Experimental Psychology: Human Perception and Performance 14*. pp. 267-280.

Moeller, S., Goedde, F., Wolters, M., 2008. Corpus analysis of spoken smart-home interactions with older users. In N. Calzolari, K.Choukri, B. Maegaard, J. Mariani, J. Odjik, S. Piperidis, and D. Tapias, (eds.), Proceedings of the Sixth International Conference on Language Resources Association. ELRA.

Rodríguez, M.D., García-Vázquez, J.P., Andrade, Á.G., 2011. Design dimensions of ambient information systems to facilitate the development of AAL environments. In *Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments*. ACM Press, New York, United States, pp. 4:1-4:7.

Roscoe, A.W., 1997. In *The Theory and Practice of Concurrency,* Prentice Hall.

Salthouse, T.A., 1994. The aging of working memory. In *Neuropsychology 8,* pp. 535-543.

Schieber, F., 1992. Aging and the senses. In *J.E. Birren, R.B. Sloane, and G.D. Cohen, (eds.) Handbook of Mental Health and Aging,* Academic Press, Volume 2.

Shi, H., Bateman, J., 2005. Developing human-robot dialogue management formally. In Proceedings of Symposium on Dialogue Modelling and Generation. Amsterdam, Netherlands.

Shi, H., Jian, C., Rachuy, C.. 2011. Evaluation of a Unified Dialogue Model for Human-Computer Interaction. In *International Journal of Computational Linguistics and Applications*. Bahri Publications, Volume 2.

Traum, D., Larsson, S., 2003. The information state approach to dialogue management. In *J.v. Kuppevelt and R. Smith (eds.), Current and New Directions in Discourse and Dialogue.* Kluwer, pp. 325-354.

Walkder, N., Philbin, D.A., Fisk, A.D., 1997. Age-related differences in movement control: adjust submovement structure to optimize performance. In *Journal of Gerontology: Psychological Sciences 52B,* pp. 40-52.

Walker, M.A., Litman, D.J., Kamm, C.A., Kamm, A.A., Abella, A., 1997. Paradise: a framework for evaluating spoken dialogue agents. In *Proceedings of the eighth conference on European chapter of Association for computational Linguistics,* NJ, USA, pp. 271-280.

Resolving Conceptual Mode Confusion with Qualitative Spatial Knowledge in Human-Robot Interaction

Cui Jian and Hui Shi

SFB/TR 8 Spatial Cognition, University of Bremen, Germany
{ken, shi}@informatik.uni-bremen.de

Abstract. This paper presents our work on using qualitative spatial knowledge to resolve conceptual mode confusion occurring frequently during the communication process between human operators and mobile robots. In order to bridge the gap between human's mental representation about space and that of a mobile robot, a qualitative spatial beliefs model is applied. Then with a computational framework based on qualitative spatial reasoning offered by this model, a set of high level strategies are developed and used to support the interpretation of natural language route instructions to mobile robots for navigation tasks.

Keywords: qualitative spatial representation and reasoning, communication of spatial information, activity-based models of spatial knowledge, human robot interaction, mode confusion.

1 Motivation and Introduction

Over the last few decades, much research has been done on intelligent robots for effectively and sensibly acting and interacting with humans in different domains. Typically, these robots are collaboratively controlled by an intelligent system and a human operator, who have to share a set of common resources such as the environment, the ongoing system's behavior and state, the remaining action plan, etc. Thus, problems may occur, when the operator's mental state about the shared common resources is different from the current system observed state, especially for the intended users of intelligent service robot, who are usually uninformed persons without specialized knowledge. These problems are called mode confusion (cf. [1]), referring to situations in which a system behaves differently from an operator's expectation. Due to its undesired consequence, mode confusion has been intensively studied, e.g. in [2], [3] and [4]. Meanwhile, with the rapid development of language technology, interaction with intelligent robots via natural language is gaining significantly increasing interest (cf. [5] and [6]), which leads to a subtype of mode confusions, called conceptual mode confusion.

Our work is focusing on resolving conceptual mode confusion occurring in human-robot joint spatially-embedded navigation tasks, where a mobile robot is instructed by a human operator via natural language to navigate in a partially known environment. Conceptual mode confusion occurs, e.g., when the human operator instructs the robot

adfa, p. 1, 2011.

to go straight ahead, take a left turn and pass a landmark on the right, but in that situation the referred landmark is only allowed to be passed after taking a right turn instead of a left turn. How this kind of spatially-related mode confusions can be detected and resolved becomes a very interesting question. [7] and [8], e.g., conducted corpus-based studies on giving route instructions to mobile robots; [9] tried to map sequences of natural language route instructions into machine readable representations; [10] reported on an approach to represent indoor environments as conceptual spatial representations with layers of map abstractions, etc.

Diverging from that literature, some research has been focusing on the perspective of human operators. [11] gave a general overview about recent reports on the human behavior and human cognition, as well as their essential relation with various spatial activities. Especially, [12] and [13] described how human thoughts and language can affect the structure of mental space for different navigation tasks. Similarly, in the scenarios of human-robot collaborative navigation, modern intelligent mobile robots have to rely on quantitative information from either pre-installed map or real-time scanners/sensors to navigate in an environment, and therefore can only accept driving requests consisting of metrical data, such as '89,45 meters ahead, then at that point make a 42,5 degree turning'. On the other hand, while interacting with mobile robots, the human operators' instructions usually contain qualitative references other than precise quantitative terms, such as "straight ahead, and then make a left turn", as well as utilizing conceptual landmarks as reorientation points in cognitive mapping (cf. [14]). There is apparently an interaction gap between a human operator and a mobile robot if either the human operator wants to send a qualitative driving command to the robot, or the robot wants to negotiate with the human about unresolved situation using its internal quantitative information. Therefore, there have been many research efforts on applying qualitative spatial calculi and models to represent spatial environments and reason about situations within those spatial settings (cf. [15], [16], [17], [18]). Adding to this body of literature, we tried to bridge the interaction gap between humans and mobile robots by introducing a qualitative spatial knowledge based intermediate level to support human-robot interaction.

In general, this paper reports our work on resolving conceptual mode confusion during the interaction process between human operators and intelligent mobile robots for spatially-embedded navigation tasks. Based on our previous work on representing and reasoning about the shared spatial related resources using a Qualitative Spatial Beliefs Model (QSBM) (cf. [19]), a computational framework is then implemented with a set of high-level strategies and used to assist human operators as well as mobile robots to detect and resolve conceptual mode confusion to interact with each other more effectively. Specifically, the first two reasoning strategies have been compared and reported in [20], the positive empirical results provided evidence on our theoretical foundation, models and frameworks. However, during the further integration within an interactive system, an additional type of conceptual mode confusions was identified and accordingly, a new high-level strategy is developed and presented in this paper.

The rest of the paper is organized as follows. We first introduce the qualitative spatial beliefs model in section 2 and a model based computational framework in section

3. Then in section 4 we present the high-level strategies, which are developed based on the QSBM model, integrated within the computational framework and used to support the resolving of conceptual mode confusion in human-robot joint navigation tasks. Then we give a conclusion and an outline to the future work in section 5.

2 Using Qualitative Spatial Knowledge to Model a Mobile Robot's Beliefs

While interacting with a mobile robot for navigation tasks, human operators often use qualitative references with vague and uncertain information to communicate with the robot. From the perspective of human operators, the navigation environment is not represented as quantitative data based map fragment as mobile robots usually do, but as a conceptual world with objects, places and the qualitative spatial relations between them. Accordingly, a qualitative spatial beliefs model is developed for representing and reasoning about spatial environments and is used to model a mobile robot's Beliefs. Then with the definition of the qualitative spatial beliefs model, we are able to define a set of update rules to support the interpretation of natural language route instructions for mobile robots in navigation tasks.

2.1 Qualitative Spatial Beliefs Model

There has been a substantial effort to develop approaches and models to represent a mobile robot's beliefs and therefore support corresponding navigation tasks. One of the widely accepted model is called Route Graph (cf. [21], [22]), which models human's topological knowledge on the cognitive level in navigation space. In conventional route graph (cf. Fig. 1 a)), all geographical places are denoted as route graph nodes with particular positions regarding a quantitative reference system, all the routes between places are then abstracted as sequences of route segments with certain lengths, directed from source nodes to target nodes. Route graphs can be used as metrical maps of mobile robots to control the navigation, because they reflect the conceptual topological structure of humans' perspective about space and therefore also ease the interaction with human to a certain extent. However, due to lack of qualitative spatial relations between places and routes, conventional quantitative route graphs are not suitable for supporting interpretation of human route instructions containing qualitative references.

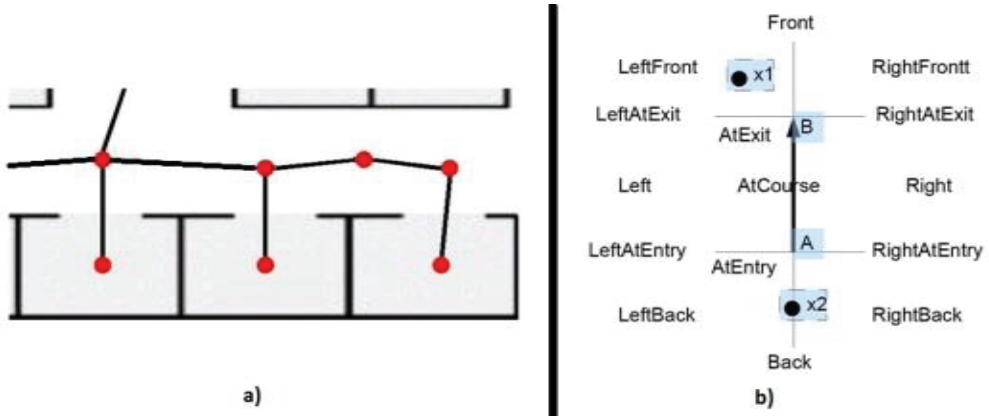


Fig. 1. A conventional route graph a) and the orientation frame of Double-Cross calculus with 15 named qualitative relations b)

Meanwhile, Double-Cross Calculus (DCC) (cf. [23]) was proposed for qualitative spatial representation and reasoning using the concept orientation grids. In this calculus, as illustrated in Fig. 1 b), a directed segment AB divides the 2-dimensional space into disjoint grids, together with the edges between the grids, 15 meaningful qualitative spatial relations (DCC relations) can be defined, such as "Front", "RightFront", "Right", etc. Thus, DCC model can describe the relative relations between objects in the local navigation map with the directed line from an egocentric perspective, e.g., the relations of x_1 and x_2 with AB can be denoted as "LeftFront" and "Back", accordingly. However, a conventional DCC model does not consider the relations between objects and places within global navigation maps, where they are connected conceptually.

Therefore, in order to benefit from both of these two models, we developed Conceptual Route Graph (abbreviated as CRG, cf. [19]), which combines the structure of conventional route graphs and the Double-Cross Calculus. On the basis of the conventional route graph and its topological representation of the places and routes in a geographical environment, qualitative spatial relations between route graph nodes and directed route segments are integrated regarding DCC relations. Thus, a conceptual route graph can be viewed as a route graph with additional qualitative DCC relations, which shows the following advantages:

- It can serve as a semantic framework for inter-process communication between different components on different levels.
- It can assist in the intuitive interpretation of human route instructions as well as appropriate presentation of internal feedback from the mobile robot system with the integrated qualitative spatial relations.
- It can be used as a direct interface with the mobile robot system for performing navigation tasks via the structure of the conventional route graph.

Given the definition of the conceptual route graph, a Qualitative Spatial Beliefs Model (abbreviated as QSBM) can be defined as a pair of a conceptual route graph

and the hypothesis of the current position of a mobile robot in the given conceptual route graph, formally denoted as <*crg, pos*>, where

- *crg* is a conceptual route graph, formally represented by a tuple of four elements (*M, P, V, R*), where
  - *M* is a set of landmarks in a spatial environment, each of which is located at a place in *P*,
  - *P* is a set of topological places on the conceptual level of the abstracted environment,
  - *V* is a set of vectors from a source place to a different target place, all of which belong to *P*,
  - *R* is a set of DCC relations between places and vectors.
- and *pos* is a directed route segment of *V*, indicating that the robot is currently located at the source place of the route segment *pos* and has the segment pointing to the target place as its orientation.

As a simple example (cf. Fig. 1 b)), let us suppose that x_1 and x_2 are two places representing a copy room and a laboratory accordingly, AB a vector from place A to place B and a mobile robot is now at the position of A looking at the place B. The qualitative spatial relations of the places x_1 and x_2 with the vector AB can be written as <*AB, LeftFront,* x_1> and <*AB, Back,* x_2>, indicating that x_1 is on the left front of AB and x_2 is at the back of AB. Therefore, a simple instance of a QSBM model of Fig. 1 b) can be specified as:

```
< crg =
        (M = {copy room:x1, laboratory:x2},
         P = {A, B, x1, x2},
         V = {AB, BA},
         R = {<AB, LeftFront, x1>, <AB, Back, x2>}),
  pos = BA >
```

2.2 Interpreting Route Instructions with Update Rules

In order to interpret natural language route instructions for navigation tasks, a mobile robot's QSBM instance should be updated according to each interpreted route instructions and provides possible feedback to the human operator. Based on the empirical studies (cf. [7], [8]) and the previous research effort concentrating on the connection between natural language, cognitive models and route instructions (cf. [9], [24], [25]), we developed a set of update rules with respect to the most common used route instructions. Formally, each update rule is defined with the following three elements:

- a name (denoted as *RULE)*, which identifies a class of human route instructions
- a set of pre-conditions (denoted as *PRE*), under which this rule can be applied, and
- an effect (denoted as *EFF*), describing how the QSBM instance should be updated after applying the update rule.

With the formal definition, the QSBM update rules are presented regarding the example classes of the common route instructions as follows:

Reorientation
defines the class of the simplest route instructions, which may change the orientation of a robot regarding the current position. "Turn left", "Turn right" and "Turn around" are the typical expressions of such route instructions. The pre-condition for Reorientation is whether the robot can find a CRG node in its QSBM instance with the following two conditions: 1. it is connected with the current position, and 2. it has the desired spatial relation with the current position; the effect is that the robot faces that found CRG node after the reorientation. Formally it is described as:

RULE: Reorientation
PRE: pos = P_0P_1,
$\exists P_0P_2 \in V$. $<P_0P_1, dir, P_2>$
EFF: pos = P_0P_2

This specifies that the robot is currently at the place P_0 and faces the place P_1 (P_0P_1 is a vector in a QSBM instance), if there exists a vector P_0P_2 in the CRG vector set V with a targeting place P_2, such that the spatial relation of P_2 with respect to the route segment P_0P_1 (or the current position) is the given direction dir to turn, i.e., $<P_0P_1$, dir, $P_2>$, then the current position will be updated with P_0P_2 after applying this update rule.

Moving Through Motion
contains a landmark, through which the robot should go, e.g., "go through the door" or "go through the lobby". These route instructions require that the given landmark is located at the extension of the current directed path.

RULE: MoveThrough
PRE: pos = P_0P_1,
$\exists P_2P_3 \in V$. $(l:P_{lm}) \wedge <P_1P_2, AtCourse, P_{lm}> \wedge$
$<P_0P_1, Front, P_2> \wedge <P_1P_2, Front, P_3>$
EFF: pos = P_2P_3

In this rule, l is the landmark given in the instruction, $(l : P_{lm})$ indicates that l is located at the place P_{lm} in the QSBM instance. The pre-condition is to find a route segment P_2P_3 in front of the current robot position P_0P_1, such that the place P_{lm} of the given landmark is located on the path of P_1P_2 (denoted as $<P_0P_1$, AtCourse, $P_{lm}>$, where AtCourse is a predefined relation in Double-Cross Calculus [23]). After applying the update rule, P_2P_3 is the new robot position.

Directed Motion
refers to the route instructions which usually contain a motion action and a turning action changing the direction of the continuing motion, such as "take the next junction

on the right". These instructions usually involve with a landmark (e.g. the "junction"), until which the robot should go, and a direction (e.g. on the "right"), towards which the robot should turn. For example, to deal with the route instruction "take the next corridor on the right", the most important step is to find the first corridor on the right from the robot's current position. Thus, the update rule for directed motions with the first landmark and a turning direction is specified as:

RULE: DirectedMotionWithFstLandmarkAndDirection
PRE: pos = P_0P_1,
$\exists P_2P_3 \in V.$ $((l{:}P_2) \wedge \langle P_1P_2, dir, P_3\rangle \wedge \langle P_0P_1, Front, P_2\rangle)$
$\wedge \forall P_4P_5 \in V.$ $((l{:}P_4) \wedge \langle P_1P_4, dir, P_5\rangle \wedge \langle P_0P_1, Front, P_4\rangle$
$\wedge (P_2 \neq P_4)) \rightarrow \langle P_1P_2, Front, P_4\rangle$
EFF: pos = P_2P_3

In this rule, l is the targeted landmark and dir is the turning direction; The first pre-condition specifies that the robot should find a route segment P_2P_3, such that the targeted landmark is located at P_2, the spatial relation between P_3 and the segment P_1P_2 before turning is the desired direction dir and P_2 is in front of the robot's current position; The second pre-condition specifies that P_2 is the first place encountered referring to the given landmark at the given direction, instead of an arbitrary one; this condition is satisfied by if there exists a place P_4 with the same feature as P_2, P_4 must be ahead of P_2 from the current perspective. Then the effect is that, the robot position is updated as P_2P_3 after applying the rule. With the definition of this rule, other variants of directed motions, such as "go straight ahead", "go right" or "take the second left" can be specified with similar update rules accordingly.

Passing Motion
refers to the route instructions containing an external landmark to be passed by, typical examples are "pass the laboratory" or "pass the laboratory on the left" with direction information. For these route instructions, the robot should first identify the landmark given in the instruction, and then check whether the landmark can be passed by along the current directed path. Furthermore, the desired direction should be considered as well, if the direction for passing the landmark is given. Accordingly the update rule PassLeft for passing a landmark on the left is specified as:

RULE: PassLeft
PRE: pos = P_0P_1,
$\exists P_2P_3 \in V.$ $(l{:}P_{lm})$
$\wedge \langle P_0P_1, LeftFront, P_{lm}\rangle \wedge \langle P_2P_3, LeftBack, P_{lm}\rangle$
$\wedge \langle P_0P_1, Front, P_2\rangle \wedge \langle P_0P_1, Front, P_3\rangle$
EFF: pos = P_2P_3

The pre-condition checks whether the current directed path satisfies the spatial requirement of the route instruction, i.e., the landmark l is located at the place P_{lm}, which is on the left front of the robot regarding the current position P_0P_1, and left behind the robot with the updated route segment P_2P_3 after executing the update rule.

Besides the above introduced update rules, there are other rules which are accordingly defined to interpret further route instructions, such as **Straight Motion**, which requests the robot to follow the current directed path; or **Moving Until Motion**, which is similar to passing motion but the robot should stop at the position parallel to the referred landmark, etc. All the QSBM update rules are implemented and integrated into a QSBM based computational framework, which is introduced in the next section.

3 A QSBM based Computational Framework

According to the introduced formal definitions of the Qualitative Spatial Beliefs Model and the QSBM update rules, we developed SimSpace, a QSBM based Computational Framework for supporting and testing the QSBM based interpretation of natural language route instructions. This section starts by introducing the architecture of the SimSpace system, and then describes how a QSBM instance can be generated from a spatial environment, and finally presents how the interpretation of the individual route instructions is supported by the SimSpace system.

3.1 System Architecture

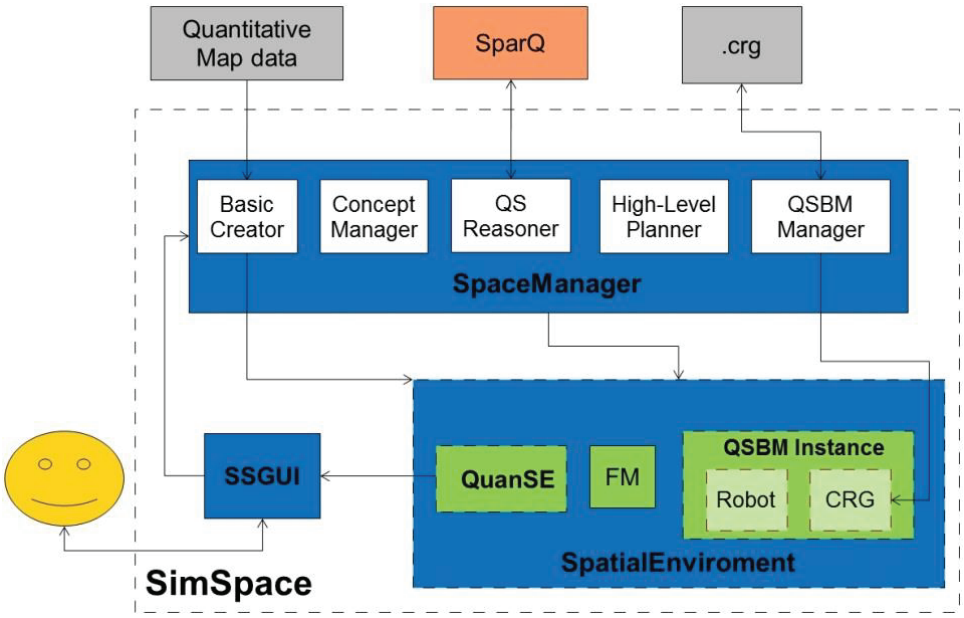


Fig. 2. The architecture of SimSpace

The architecture of the SimSpace system is illustrated in Fig. 2. It consists of two major components and one optional component, together with the external resources presented as follows:

- **The external resources** include

  - the quantitative map data containing quantitative and conceptual information about a certain spatial environment,
  - the conceptual route graph file (.crg), which is a XML-based specification of a the conceptual route graph of a QSBM instance, and
  - the qualitative spatial knowledge based toolkit SparQ (cf. [26]), connected with the SimSpace system to support the qualitative spatial representation and reasoning about the QSBM instance of the spatial environment.
- **The component Spatial Environment** maintains the QSBM instance with the conceptual route graph and the hypothesis of the robot position in the CRG as defined before, as well as the optional quantitative spatial environment (QuanSE) for quantitative data and the optional feature map (FM) component containing the conceptual information of the environment.
- **The processing component Space Manager** is the central processing unit of the SimSpace system with the following functional components:
  - Basic Creator creates a spatial environment instance with quantitative and conceptual data according to the quantitative map data, if given.
  - Concept Manager manages an ontology-based database of the none-spatial conceptual knowledge, such as names of locations or persons, how they are conceptually related, etc. It is used to interpret the conceptual terms in the natural language route instructions.
  - QS Reasoner is responsible for the direct communication with the SparQ toolbox and handles the most basic operations of qualitative spatial representation and reasoning in QSBM.
  - High-Level Planner integrates all the QSBM update rules and implements a set of high-level strategies to choose and apply appropriate update rules to interpret route instructions and resolve conceptual mode confusion. The planning process is detailed in the next section.
  - QSBM Manager generates a QSBM instance according to a quantitative environment if given, manipulates and updates an existing QSBM instance, and saves it into a XML-based specification, if needed.
- **The interaction component SSGUI** is the graphical user interface of SimSpace. It is an optional component and is only used if the SimSpace system is started as a stand-alone application. The current SSGUI visualizes the spatial environment with quantitative and conceptual descriptions, interacts with a human user who is giving the natural language route instructions, and communicates with the Space Manager for the interpretation of incoming route instructions as well as outgoing system responses.

3.2 Construction of a QSBM instance

In order to support the QSBM based interpretation of route instructions, a conceptual route graph in QSBM regarding a specific spatial environment needs to be constructed. One possible way is to use existing quantitative data. SimSpace takes a conventional quantitative route graph as input and constructs a corresponding DCC-based conceptual route graph in two steps:

- The quantitative data is qualified into DCC relations with the qualify module of the SparQ toolkit.
- The DCC relations cannot be used directly, instead they need to be generalized, i.e., the relations regarding angles near 0, 90 and 180 are accordingly assigned to those matching exactly 0, 90 and180.

Despite the fact that a slight loss of precision may occur, the generalization is necessary. First and foremost, a conceptual route graph serves as an interface with human operators for navigation tasks, where they usually generalize the perceived qualitative relations mentally; therefore human operators usually use the generalized relations in the route instructions (cf. [27]). Instead of saying 'take a LeftFront turn', e.g., they usually say 'take a left turn'. Moreover, the qualitative spatial reasoning with ungeneralized relations provides too many possible results even after one step calculation, which is unlikely to handle.

A conceptual route graph of a QSBM instance can also be generated dynamically, if it is used for a mobile robot navigating in an unknown or partially known environment. The conceptual route graph is empty or with the partially known graph connections at the beginning, and keeps being updated by the QSBM manager while performing the navigation tasks, either via the collaborative interaction with the human operator, or with the real-time sensory data of the mobile robot.

3.3 Interpreting route instructions with SimSpace

With a QSBM instance generated, SimSpace can interpret a human route instruction in the following steps:

- The given natural language route instruction is firstly parsed into a pre-defined semantic representation.
- According to the category of the semantic representation, an applicable QSBM update rule is chosen and its pre-conditions are instantiated. Taking the sample instruction "pass the laboratory on the left" in the previous section, the update rule PassLeft is applied and by assuming AB to be the current robot position, the second pre-condition is instantiated to:

$$\exists P_2P_3 \in V.\ (\text{Laboratory}:P_{lab}) \wedge \langle AB, \text{LeftFront}, P_{lab}\rangle \wedge \langle P_2P_3, \text{LeftBack}, P_{lab}\rangle \wedge \langle AB, \text{Front}, P_2\rangle \wedge \langle AB, \text{Front}, P_3\rangle$$

- Then with the SparQ toolkit, the instantiated pre-conditions are checked against the current state of the QSBM instance. If the current state is matched with the instantiated pre-condition, the current robot position is updated to P_2P_3; If however the current state provides e.g. the following relations:

$$(\text{Laboratory}:P_{lab}) \wedge \langle AB, \text{RightFront}, P_{lab}\rangle$$

i.e., the laboratory is located on the right side regarding the perspective of the current robot position, which means that $\langle AB, \text{LeftFront}, P_{lab}\rangle$ in the pre-condition

cannot be satisfied. In this case, SimSpace interprets these results into a corresponding representation to allow the generation of a clarification to the human operator.

Generally, the SimSpace system combines the implementation of the QSBM and the update rules. As a well encapsulated module, it can be integrated into an interactive mobile robot system to assist in the interaction with human operators via its intuitive qualitative spatial representation and reasoning about the spatial environment and set up a direct communication with the mobile robots via the inherited features from conventional route graphs. It can also be used as a stand-alone evaluation platform for visualizing spatial environments, generating corresponding QSBM instances and testing the interpretation of natural language route instructions. However, in order to interpret a sequence of natural language route instructions and resolve the possible consequent conceptual mode confusions, a set of high-level strategies are needed, which are developed and introduced in the following section.

4 Resolving Conceptual Mode Confusion with High-Level Strategies

For human operators, giving a sequence of route instructions to mobile robots may cause conceptual mode confusions, because before they can organize the appropriate terms for the route instructions, they first need to correctly locate the robot's current position and the desired goal location, and then take the imagined journey in mind to go along the expected route, while working in possible mental rotation during the travelling. Due to this complicated process, a wrongly located place or taken turn, which happens quite often (cf. [28]), would cause the failure of the interpretation of the entire route instructions and consequently lead to conceptual mode confusion situations. In order to cope with these problems, a set of high-level strategies based on QSBM and the QSBM update rules are developed and presented in this section.

4.1 Deep Reasoning

One of the most typical conceptual mode confusions is spatial relation or orientation mismatches (cf. [29]). This type of conceptual mode confusions occurs, if a spatial object is incorrectly orientated in the operator's mental representation, such as the previous example "pass the laboratory on the left", where the laboratory can only be passed on the right; or "take the next junction on the right", where the next junction is only leading to the left.

Given a QSBM instance and the appropriate QSBM update rule, the expected state of route instructions will be checked against the actual QSBM observed state using qualitative spatial reasoning. If an unsatisfied situation is identified, it can be presented to the human operator appropriately. If possible, while checking the instantiation of the route instruction leading to the unsatisfied situation, a corrected spatial or ori-

entation relation is inferred for resolving the confusion. Therefore, the deep reasoning strategy tries to resolve the problematic situation by either giving a reason regarding the current situation to support the human operator to reorganize the route instructions, such as "you can't pass the laboratory on the left, because it is now behind you". More intuitively, it can make a suitable suggestion if existed, such as "you can't take a right turn here, but maybe you mean to take a left turn?"

4.2 Deep Reasoning with Backtracking

Besides the straightforward conceptual mode confusions concerning with the mentally wrongly oriented objects, there are situations where route instructions involve mental travelling or rotation that could easily be made incorrectly. If one instruction is wrongly given, the rest of the route instructions cannot be interpreted because the actual state caused by the wrong instruction does not match the mental state of the operator. Fig. 3 illustrates a typical example of this type of conceptual mode confusions, where the position of the mobile robot is shown by the arrow and the sequence of route instructions is "go straight, go left, left again, pass by B1 on the right". Using the deep reasoning strategy in the previous subsection, the robot will go along the path following the first three instructions, and then have a problem to interpret the last one "pass by B1 on the right". Finally it can only provide a reason why the last instruction cannot be taken, like "you can't pass by B1, because it's now behind you". However, by taking one step backwards, instead of "turn left", if the instruction is "turn right", then the last instruction can be interpreted appropriately.

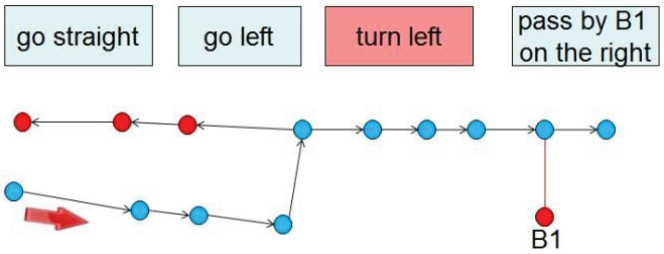


Fig. 3. An example for the conceptual mode confusions caused by an incorrect instruction

Therefore, the strategy "Deep Reasoning with Backtracking" manages the application of QSBM update rules with respect to the matching route instructions as "Deep Reasoning" usually does. Moreover, after applying one update rule for each route instruction, the state of the updated QSBM instance is kept in a transition history. If the checking of the current system state against a route instruction fails, the previous state is reloaded from the transition history. Based on it, possible correction/suggestion can be made, such as "turn right" substituting "turn left" in the example of Fig. 3, so that the interpretation of the remaining route instructions can proceed. In this case, instead of giving a reason regarding the first encountered uninterpretable route instruction, deep reasoning with backtracking tries to locate the potentially wrongly made route

instructions in the interpretation history, then resume the checking of the current route instruction with a possible correction/suggestion, and finally finds a successful interpretation if one exists.

4.3 Searching with QSR-weighted value tuples

Although deep reasoning with backtracking interprets more route instructions and better supports resolving conceptual mode confusion compared to the deep reasoning strategy (cf. [19] and [20]), an additional type of conceptual mode confusions regarding incorrectly located starting or turning position (called "conceptual location mismatch") is found. Fig. 4 illustrates two examples of conceptual location mismatches.

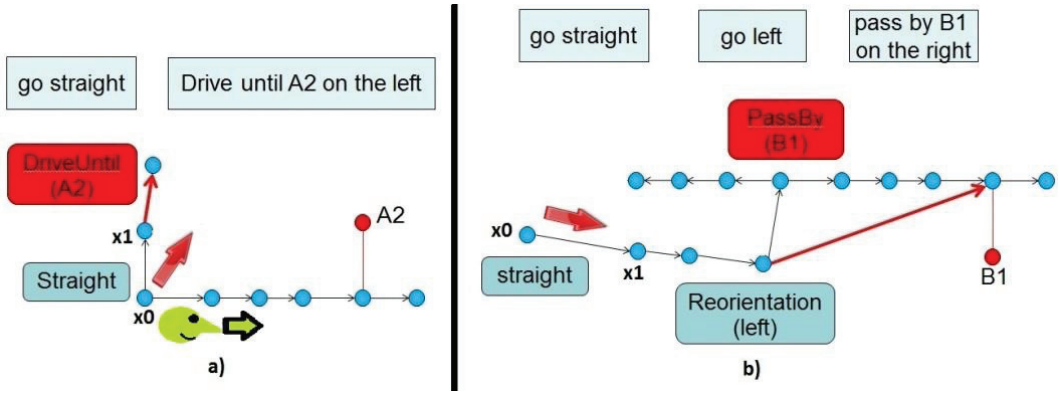


Fig. 4. Two examples of conceptual location mismatch

In example of Fig. 4 a), the position of the mobile robot is shown by the thick arrow, slightly nearer to the path leading upwards, and therefore the hypothesized robot position in the QSBM instance is represented by the route segment x0x1 leading upwards. However, the human operator, who is looking at the right direction in the illustrated map, thinks the robot is looking at the same direction and gives the instruction: "straight ahead and drive until A2 on the left". In the example of Fig. 4 b) with the robot position x0x1 and the route instructions "go straight and go left and pass by B1 on the right", after taking left, the operator thinks B1 is now located directly on the right from her/his perspective, and therefore simply ignores a turning point in the conceptual representation of the QSBM instance.

Both deep reasoning and deep reasoning with backtracking cannot provide an appropriate solution for resolving this type of conceptual mode confusions, because suggestions can only be made according to the given route instructions, while in such cases one instruction is missing and no possible suggestion can be made. Thus, the additional strategy of searching with Qualitative-Spatial-Relation-weighted (abbreviated as QSR-weighted) value tuples is developed.

First, we defined a set of QSR weighted value tuples for each starting/turning/decision point with every outgoing direction as:

```
{(route, instructions, qsr-v)*}
```

Here "route" represents the currently taken route, "instructions" includes all the along this route interpreted instructions, and "qsr-v" is the cumulative value calculated by

$$\sum (mr_i * sr_i)$$

where mr_i is the matching rate by comparing the taken qualitative spatial direction with the current route direction at the i-th decision point, and sr_i is the success rate of interpreting a route instruction at that point.

With the definition of QSR-weighted value tuples, finding an appropriate interpretation (namely a route) to correspond to a sequence of natural language route instructions is illustrated in Fig. 5. The QSBM manager first initializes an empty set of QSR-weighted value tuples at the current position of the robot (the black point in the middle of the network in Fig. 5). This value-tuple-set is then automatically updated by the QSBM manager (as {(r1, i1, v1), (r2, i2, v2), (r3, i3, v3)} in Fig. 5, where (rx, ix, vx) indicates the tuple of the covered route rx, the interpreted instructions ix and the relating QSR-weighted value vx). Searching agents of the QSBM manager are then travelling along all paths (according to the branching of the current point, e.g. three paths in Fig. 5) on the current QSBM. The value-tuple-set gets updated and expanded by the QSBM based update rules when new branches are encountered or new instructions are interpreted. Finally, a full set of value-tuples is generated. The value tuple with the highest QSR-weighted value is either the best possible solution for interpreting the given route instructions or contains the most relevant information to provide the possible suggestion/correction to resolve the conceptual mode confusions.

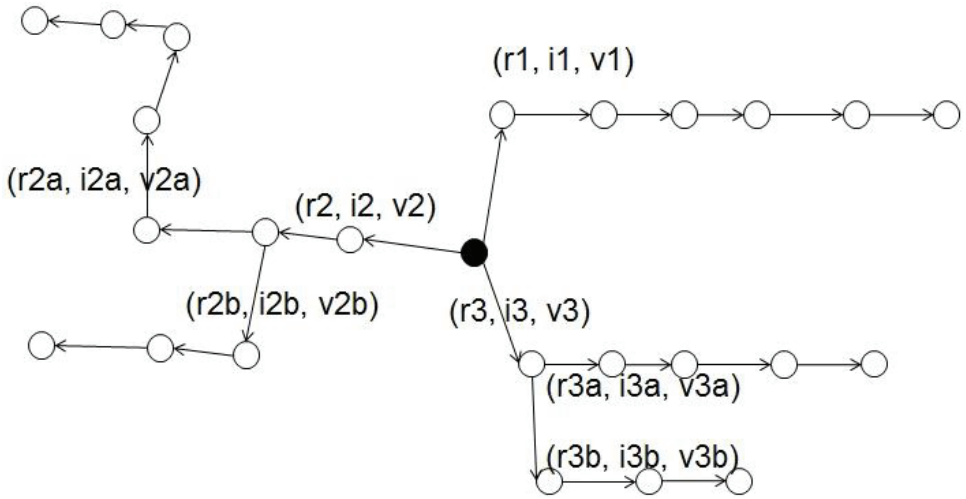


Fig. 5. The searching with QSR-weighted value tuples

With this strategy, the two example situations caused by the conceptual location mismatches in Fig. 4 can be solved as illustrated in Fig. 6.

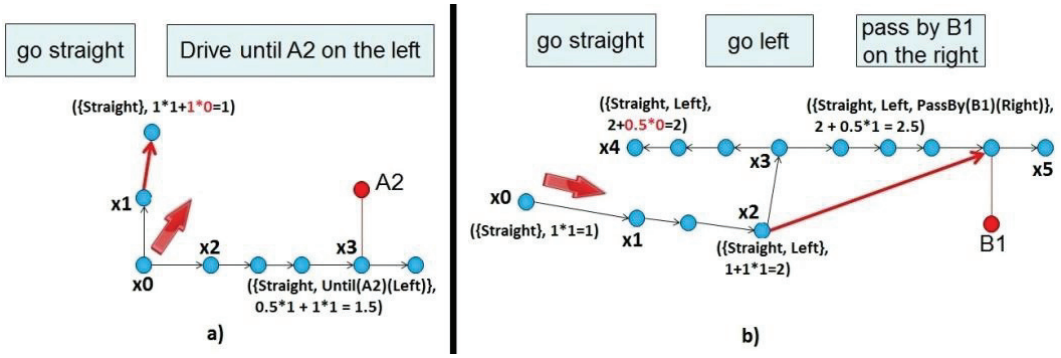


Fig. 6. Illustrated processes of solving conceptual location mismatch using QSR-Weighted value tuples (the route element in each QSR-weighted value tuple is ignored for simplicity)

In the first example of Fig.6 a) with the starting place x0, the set of the QSR-weighted value tuples is:

```
{(x0->x1, {Straight}, 1   * 1  = 1),
 (x0->x2, {Straight}, 0.5 * 1 = 0.5)}
```

Here for the route x0->x1, the matching rate is 1 because the route matches exactly the starting segment x0x1, the success rate of interpreting "go straight" is 1, and therefore the QSR-weighted value is 1*1 = 1. For the route x0->x2, the qualitative relation of the place x2 with the starting segment x0x1 is RightAtExit, so the matching rate is assigned as 0.5, however the success rate is again 1, so the QSR-weighted value is 0.5*1 = 0.5. Similarly, as the searching goes on, the route starting from x0x1 keeps going straight with the matching rate 1, but fails to interpret the instruction "Drive until A2 on the left" and gets the success rate 0; while the other route from x0x2 is also leading straight as well as being able to interpret the second instruction. Therefore, the instructions are interpreted with the route starting from x0x2 and a higher QSR-weighted value 1.5.

In the second example of Fig. 6 b), the searching goes along one route while successfully interpreting the first two instructions "go straight, and then left" with the following QSR-weighted value tuple:

```
(x0->x1-x2->x3, {Straight, Left}, 1*1 + 1*1 = 2)
```

When the searching comes to the turning point x3 with the instruction "pass by B1 on the right", there are two routes going into the left and right directions with the matching rates 0.5, but the route leading to the left cannot interpret the last instruction while the other one can. Therefore, the instructions are interpreted with the route x0->x1->x2->x3->x5, since it has the higher QSR-weighted value 2.5.

On the one hand, the strategy of searching with QSR-weighted value tuples resolves conceptual mode confusion from the perspective of a mapping problem in a directed graph with QSR weighted values. On the other hand, it preserves the functionality of deep reasoning with the QSBM update rules and the QSBM instance on each outgoing path from each decision point. Therefore, it can be viewed as a search-

ing algorithm with multiple deep reasoning agents supporting the interpretation of more route instructions while clarifying more conceptual mode confusions.

5 Conclusion and Future Work

This paper has presented our research work on three important aspects of Cognitive Science: representing and managing, and reasoning with, qualitative spatial knowledge. Specifically, we focus on resolving conceptual mode confusion during the natural language based interaction between a human operator and a mobile robot for spatially-embedded navigation tasks. In order to support effective and intuitive human robot interaction, a qualitative spatial beliefs model has been developed for representing the shared conceptual spatial environment and a model-based computational framework has been accordingly implemented. Together with a set of high-level QSBM-based strategies, especially the strategy of searching with qualitative spatial relation weighted value tuples, mobile robots can be assisted to detect and resolve different types of conceptual mode confusions for more effective and intuitive interaction with human operators.

The reported work continued the pursuit of our goal towards building effective, intuitive and robust interactive frameworks and systems in spatially-related settings. Currently, we are integrating the qualitative spatial beliefs model and the model-based computational framework into a natural language interactive dialogue system for navigating a mobile robot in indoor environments. Regarding the high-level strategies, especially how to combine and apply the deep reasoning with backtracking and searching with QSR-weighted value tuples to better support natural language based spatially-related interaction is being investigated with empirical studies. Further qualitative spatial calculi and models are also being considered to extend the current qualitative spatial beliefs model for supporting more application domains. Human-robot collaborative exploration and navigation in unknown or partially known spatial environment is also another work package to be covered in the next stage.

Acknowledgement

We gratefully acknowledge the support of the German Research Foundation (DFG) through the Collaborative Research Center SFB/TR 8 Spatial Cognition - Subproject I5- [DiaSpace].

References

1. Sarter, N., Woods, D.: How in The World Did We Ever Get into That Mode? Mode error and Awareness in Supervisory Control. Human Factors 37, 5-19 (1995)
2. Bredereke, J., Lankenau, A.: A Rigorous View of Mode Confusion. In: Proceeding of Safecomp 2002, 21st International Conference on Computer Safety, Reliability and Security. LNCS, vol. 2434, pp. 19–31. Springer Verlag, London, UK (2002)

3. Lüttgen, G., Carreno, V.: Analyzing Mode Confusion Via Model Checking. Technical Report, Institute for Computer Applications in Science and Engineering (ICASE) (1999)
4. Heymann, M., Degani, A.: Constructing Human-Automation Interfaces: A Formal Approach. In: Proceedings HCI-Aero 2002, pp. 119-125. Cambridge, MA (2002)
5. Augusto, J. C., McCullagh, P.: Ambient intelligence: Concepts and Applications. Computer Science and Information Systems, vol. 4(1), pp. 228–250 (2007)
6. Montoro, G., Haya, P.A., Alamán, X.: A Dynamic Spoken Dialogue Interface for Ambient Intelligence Interaction. In International Journal of Ambient Computing and Intelligence, Vol. 2(1), pp. 24-51, IGI Publishing Hershey, PA, USA (2010)
7. Bugmann, G., Klein, E., Lauria, S., Kyriacou, T.: Corpus-Based Robotics: A Route Instruction Example. In Proceedings of the 8th International Conference on Intelligent Autonomous Systems (IAS-8), pp. 96-103, Amsterdam, Netherlands (2004)
8. Shi, H., Tenbrink, T.: Telling Rolland Where to Go: HRI Dialogues on Route Navigation. In: Conventry, K., Tenbrink, T., Bateman, J. (eds.) Spatial Language and Dialogue, pp. 177–190, Cambridge University Press (2009)
9. Lauria, S., Kyriacou, S., Bugmann, G., Bos, J., Klein, E.: Converting Natural Language Route Instructions into Robot Executable Procedures. In: Proceedings of the 2002 IEEE International Workshop on Human and Robot Interactive Communication, pp. 223–228 (2002)
10. Zender, H., Mozos, O.M., Jensfelt, P., Kruijff, G.-J.M., Burgard, W.: Conceptual Spatial Representations for Indoor Mobile Robots. In International Journal of Robotics and Autonomous Systems. Vol. 56(6), pp. 493-502, Amsterdam, Netherlands (2008)
11. Denis, M., Loomis, J. M.: Perspectives on Human Spatial Cognition: Memory, Navigation, and Environmental learning. In Psychological Research. Vol. 71(3), pp. 235-239, Springer-Verlag (2007)
12. Tversky, B., Kim, J.: Mental Models of Spatial Relations and Transformations from Language. In: Habel, C., Rickheit, G. (eds.) Mental Models in Discourse Processing and Reasoning, pp. 239-258 (1999)
13. Tversky, B.: Structures of Mental Spaces: How People Think About Space. In: Environment & Behavior, pp. 66-80 (2003)
14. Michon, P.E., Denis, M.: When and Why Are Visual Landmarks Used in Giving Directions. In: Proceedings of the International Conference on Spatial Information Theory: Foundations of Geographic Information Science, pp. 292-305 (2001)
15. Skubic, M.: Qualitative Spatial Referencing for Natural Human-Robot Interfaces. In: interactions. Vol. 12(2), pp. 27-30, ACM New York, USA (2005)
16. Wallgrün, J.O., Wolter, D., Richter, K-F.: Qualitative Matching of Spatial Information. In: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '10), pp. 300-309, ACM, New York, USA (2010)
17. Kurata, Y.: 9+-Intersection Calculi for Spatial Reasoning on the Topological Relations between Heterogeneous Objects. In: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '10), pp. 390-393, ACM, New York, USA (2010)
18. Liu, W.M., Wang, S.S., Li, S.J., Liu, D.Y.: Solving Qualitative Constraints Involving Landmarks. In: Lee, J. (eds.) Proceedings of the 17th international conference on Principles and practice of constraint programming (CP'11), pp. 523-537, Springer-Verlag, Berlin, Heidelberg (2011)
19. Shi, H., Jian, C., Krieg-Brückner, B.: Qualitative Spatial Modelling of Human Route Instructions to Mobile Robots. In Proceedings of the 2010 Third International Conference on

Advances in Computer-Human Interactions (ACHI '10), pp. 1-6, IEEE Computer Society, Washington, DC, USA (2010)
20. Jian, C., Zhekova, D., Shi, H., Bateman, J.: Deep Reasoning in Clarification Dialogues with Mobile Robots. In: Coelho, H., Studer, R., Wooldridge, M. (eds.) Proceedings of the 2010 conference on 19th European Conference on Artificial Intelligence (ECAI), pp. 177-182, IOS Press, Amsterdam, The Netherlands (2010)
21. Werner, S., Krieg-Brückner, B., Herrmann, T.: Modelling Navigational Knowledge by Route Graphs. In: Freksa, C., Brauer, W., Habel, C., Wender, K.F. (eds.) Spatial Cognition II, Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications, pp. 295-316, Springer-Verlag, London, UK (2000)
22. Krieg-Brückner, B., Frese, U., Lüttich, K., Mandel, C., Mossakowski, T., Ross, R.J.: Specification of an Ontology for Route Graphs. In: Freksa, C., Knauff, M., Krieg-Brückner, B., Nebel, B., Barkowsky, T. (eds.) Proceedings of Spatial Cognition IV. Lecture Notes in Artificial Intelligence, Vol. 3343, pp. 989–995, Springer, Chiemsee, Germany (2004)
23. Freksa, C.: Using Orientation Information for Qualitative Spatial Reasoning. In Theories and Methods of Spatio-Temporal Reasoning in Geographic Space. Lecture Notes in Computer Science, Vol. 639, pp. 162–178, Springer-Verlag (1992)
24. Denis, M.: The Description of Routes: A Cognitive Approach to the Production of Spatial Discourse. In: Cahiers de Psychologie Cognitive, 16, pp. 409–458 (1997)
25. Tversky, B.,Lee, P.U.: How Space Structures Language. In: Freksa, C., Habel, C., Wender, K.F. (eds.) Spatial Cognition: An interdisciplinary Approach to Representation and Processing of Spatial Knowledge. Lecture Notes in Artificial Intelligence, Vol. 1404, pp. 157–175, Springer-Verlag London, UK (1998)
26. Wallgrün, J.O., Frommberger, L., Wolter, D., Dylla, F., Freksa, C.: Qualitative Spatial Representation and Reasoning in the SparQ-Toolbox. In: Barkowsky, T., Knauff, M., Ligozat, G., Montello, D.R. (eds.) Proceedings of the 2006 International Conference on Spatial Cognition V: Reasoning, Action, Interaction. Lecture Notes in Computer Science, Vol. 4387, pp. 39–58, Springer-Verlag, Berlin, Heidelberg (2007).
27. Montello, D. R.: Spatial Orientation and the Angularity of Urban Routes — A Field Study. In: Environment and Behavior, Vol. 23(1), pp. 47–69 (1991)
28. Reason, J.: Human Error. Cambridge University Press (1990)
29. Shi, H., Krieg-Brückner, B.: Modelling Route Instructions for Robust Human-Robot Interaction on Navigation Tasks. In: International Journal of Software and Informatics, vol. 2(1), pp. 33–60 (2008)

Modality Preference in Multimodal Interaction for Elderly Persons

Cui Jian[1], Hui Shi[1], Nadine Sasse[2], Carsten Rachuy[1], Frank Schafmeister[2], Holger Schmidt[3], and Nicole von Steinbüchel[2]

[1]SFB/TR8 Spatial Cognition, Universität Bremen, Germany
{ken, shi, rachuy}@informatik.uni-bremen.de
[2]Medical Psychology and Medical Sociology, University Medical Center Göttingen, Germany
{n.sasse, frank-schafmeister, nvsteinbuechel}@med.uni-goettingen.de
[3]Neurology, University Medical Center Göttingen, Germany
h.schmidt@med.uni-goettingen.de

Abstract. This paper is focusing on two important aspects: on the one hand, it presents our work on designing, developing and implementing a multimodal interactive guidance system for elderly persons to be used in autonomous navigation within complex building; on the other hand, it summaries and compares the data of a series of empirical studies that have been conducted to evaluate the effectiveness, efficiency and user satisfaction of the elderly-centered multimodal interactive system regarding different multimodal input-possibilities such as speech, gesture via touch-screen and the combination of both under simulated conditions. The overall positive results validated our systematically developed and empirically improved design guidelines, foundations, models and frameworks for supporting multimodal interaction for elderly persons.

Keywords: Multimodal interaction, elderly-friendly interface, dialogue management, human-computer interaction in AAL, formal methods, interactive system evaluation

1 Introduction

There has been a rapidly growing interest in research and development of multimodal interaction over the past few decades ([1] and [2]). Specifically, multimodal interaction is showing its importance and necessity for more effective interaction compared to single modal interaction (see e.g. [3] and [4]). It also holds the potential of further enhancing users' individual as well as overall performance (see e.g. [5]). Moreover, multimodal interfaces can be used to achieve a more natural human-computer communication and increase the robustness of the interaction with complementary information (see e.g. [6]).

However, the typical multimodal interaction mechanisms are usually only suitable for users with sufficient familiarity with information and communication technology, which poses a particular challenge for people with less knowledge about this kind of

adfa, p. 1, 2011.

interaction, especially for the constantly growing group of elderly persons due to the acceleration of population ageing nowadays in almost all industrialized countries ([7]). Therefore, in order to maximize the advantage of multimodal interaction, special focus has been laid on the research of multimodal interaction with respect to the emerging area Ambient Assisted Living and its potential user group: elderly persons or persons with special needs (see e.g. [8], [9] and [10]).

Adding to this body of literature, our work is concentrating on multimodal interaction in AAL context for elderly persons by taking ageing-related characteristics into account. It can be divided into two important aspects: a) the design, development and implementation of multimodal interaction for elderly persons; and b) the empirical evaluation of a minutely developed and systematically improved elderly-friendly multimodal modal interactive system with elderly persons.

For a) two fundamental aspects are proposed for supporting our system design and development: a list of elaborated design guidelines regarding traditional design principles of conventional interactive systems and the most common elderly-centered characteristics corresponding to ageing-related decline of sensory, perceptual, motor and cognitive abilities of elderly persons ([11]); and a formal language supported unified dialogue modelling and management approach, which combines a finite state based generalized dialogue model and a classic agent based management theory, and therefore can support a flexible and context-sensitive, yet formally tractable and controllable multimodal interaction ([12]). According to the two development foundations, a multimodal interactive guidance system was then especially designed and implemented.

For b) a series of empirical studies were conducted with groups of elderly persons to practically evaluate the multimodal interactive system with respect to its multiple input modalities as well as to enable a continuously improved development based on the data of each empirical study. Specifically, a touch-screen graphical user interface was implemented and tested in a pre-study in [11]; a spoken language interface and its dialogic interaction were tested in [12]; with the test data, the touch-screen interface and the spoken language interface are accordingly improved, tested and compared with each other in a follow-up study in [13]; then the combination of the two modalities are evaluated in the next study and the results are described in [14].

Therefore, in order to perform a detailed comparison of all the input modalities, namely the touch-screen, the spoken language and the combination of both, as well as the assessment of the complete multimodal interactive system concerning its effectiveness of task performance, efficiency of interaction and user satisfaction about the system, the data of all the experimental studies were summarized and analyzed, then the results are described and discussed in this paper.

The rest of the paper is structured as follows: section 2 briefly introduces the proposed and improved design and development foundation of our work and presents the minutely implemented multimodal interactive system for elderly persons; section 3 describes the experimental studies on evaluating the complete interactive system while focusing on comparing the multiple input modalities; section 4 summarizes and analyzes the empirical data and discussed about the results according to an adapted version of a classic evaluation framework for conventional interactive systems; final-

ly, section 5 gives a conclusion of the reported work and outlines the direction of our future research focus.

2 System Design, Development and Implementation

This section first introduces our theoretical foundation for designing and developing multimodal interaction for elderly persons, then according to the design and development foundation a multimodal interactive guidance system for elderly persons is implemented and presented.

2.1 The Foundation of System Design and Development

The theoretical foundation of our work consists of two aspects: a set of guidelines to support the design and development of elderly-friendly multimodal interaction; and a unified dialogue modelling approach with a formal method based framework for dialogue management.

Design Guidelines of Multimodal Interaction for Elderly Persons.

During the ageing process, elderly persons often suffer from decline of sensory, perceptual, motor and cognitive abilities, especially for the seven most common human abilities, as shown in Fig. 1.

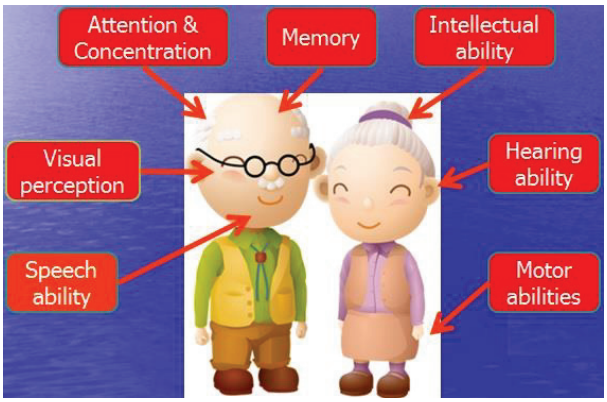


Fig. 1. The seven most common ageing-related human abilities

Specifically, **visual perception** declines for most people with ageing; physically the size of the visual field is decreasing and the peripheral vision can be lost; It is more difficult to focus on objects up close and to see fine details, including rich colors and complex shapes that make images hard or even impossible to identify; rapidly moving objects are either causing too much distraction, or become less noticeable ([15]); **speech ability** declines while ageing in the way of being less efficient for pronouncing complex words or longer sentences, probably due to reduced motor control of tongue or lips ([16]); [17] also confirmed that, elderly-centered adaptation of speech-

enabled interactive components can improve the interaction quality to a satisfactory level; **attention and concentration** drop while ageing, elderly persons either become more easily distracted by details and noise, or find other things harder to notice when concentrating on one thing ([18]); they show great difficulty with situations where divided attention is needed ([19]); **memory functions** decline differently. Short term memory holds fewer items with age and working memory becomes less efficient ([20]). Semantic information is normally preserved in long term memory ([21]); **intellectual ability** does not decline much during the normal ageing process, yet [22] believed that crystallized intelligence can assist elderly persons to perform better in a stable well-known interface environment; **hearing ability** declines to 75% between the age of 75 and 79 ([23]). High pitched sounds are hard to perceive; complex sentences are difficult to follow ([24]); **motor abilities** decline generally due to loss of physical activities while ageing. Complex fine motor activities are more difficult to perform, e.g. to grab small or irregular targets ([25]); conventional input devices such as a computer mouse are less preferred by elderly persons as good hand-eye coordination is required ([26]).

According to the above empirical findings and much more other research work on effects of ageing using computer based systems (see e.g. [27], [28], [29]), it is necessary to consider age-related characteristics while developing interactive systems for elderly persons. Therefore, based on the common design principles for conventional interactive systems and the ageing-related characteristics regarding the seven most common human abilities, a set of guidelines for designing and developing multimodal interactive system for elderly persons was proposed in [11]. These guidelines have been implemented into the first versions of our interactive systems, evaluated by our previous empirical studies with elderly persons, and then accordingly improved on the basis of the evaluation results. The final set of improved design guidelines were summarized in [13] and have been used as the first fundamental aspect of our work ever since, especially for the development of the final version of our multimodal interactive system.

Formal Language Supported Unified Dialogue Modelling and Management.

One of the most essential issues of developing an interactive system is the interaction management, i.e., how the interaction flow is controlled in the dialogues between users and the system. In most of the related work on dialogue modelling and management, the following two methods can be deemed as the basic and important ones among others:

- The generalized dialogue models, which can abstract dialogue models by describing discourse patterns as illocutionary acts in the classic recursive transition networks, without reference to any direct surface indicators ([30]);
- The information state update based management theories, which focus on the modelling of discourse context as the attitudinal state of an intelligent agent and show a powerful way to handle dynamic information for a context sensitive dialogue management ([31]).

However, these two methods have their own drawbacks. On the one hand, the generalized dialogue models are based on finite state transition models, which are criticized for their inflexibility of dealing with dynamic information; on the other hand, the information state update models have the problem of controlling their complexity for state manage and model extension.

Therefore, a unified dialogue modelling approach is proposed ([11]), which extends a generalized dialogue model with the information state update based components, such that finite state transitions can only be triggered by fulfilled conditions and followed by updated information state with a set of predefined information state update rules (see the left part of Fig. 2).

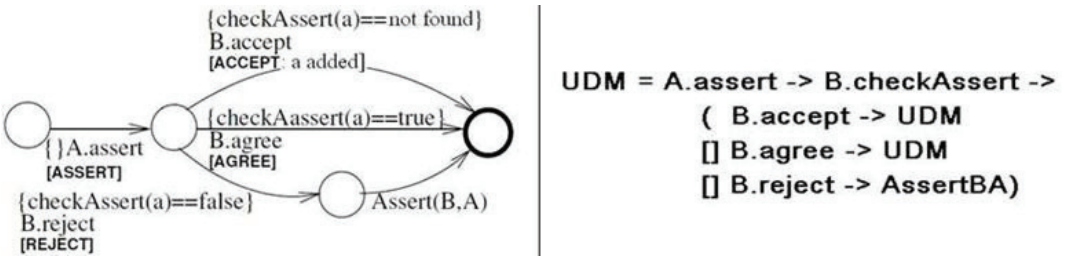


Fig. 2. A simple unified dialogue model with its CSP Specification on the illocutionary level

In order to support the development and implementation of unified dialogue models and their integration into practical interactive systems, the formal dialogue development framework (abbreviated as FormDia) was developed and proposed in [32].

Fig. 3 illustrates the structure of the FormDia framework, which consists of six important components according to the development process of a unified dialogue model based dialogue manager:

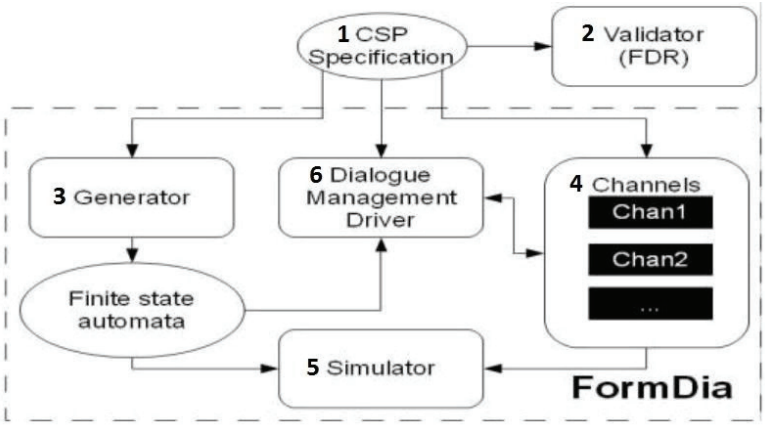


Fig. 3. The structure of the FormDia framework

1. **CSP Specification**: every unified dialogue model is based on a generalized dialogue model, whose illocutionary structure can therefore be specified as a machine

readable Communication Sequential Processes (CSP [33]) program (see an example CSP specification of a simple unified dialogue model in the right part of Fig. 2)
2. **Validator**: the CSP specified program can then be validated by the Failures-Divergence Refinement tool, (FDR [34]), which is a model checking tool for validating and verifying the properties of CSP specifications.
3. **Generator**: according to the validated CSP specification, machine readable finite state automata can be generated by the Generator.
4. **Channels**: with the finite state automata, channels regarding all the generated states can be defined with related information state update rules. These channels are at first black boxes, which will be filled with deterministic behavior of concrete components according to their application contexts.
5. **Simulator**: with the finite state automata and the defined communication channels, dialogue scenarios can be simulated via a graphical interface, which visualizes dialogue states as a directed graph and provides a set of utilities to trigger dialogue events and updating of dialogue states for testing and verification.
6. **Dialogue Management Driver**: after validation and verification, the unified dialogue model can be integrated into a practical interactive dialogue system via the dialogue management driver.

The unified dialogue modelling approach with the formal language based dialogue management framework FormDia is detailed in [11] and serves as the second fundamental aspect of our work. A unified dialogue model for the multimodal interaction for elderly persons was developed. With the FormDia framework, this model is implemented and integrated into our interactive system for elderly persons, then evaluated via empirical studies and improved accordingly (see e.g. [12], [13]).

2.2 System Description

Fig. 4. An autonomous electronic wheelchair ([35]) to be interacted via MIGSEP with an elderly or handicapped person

According to the two introduced development foundations, we developed MIGSEP, a general Multimodal Interactive Guidance System for Elderly Persons. MIGSEP runs on a portable touch-screen tablet PC and will serve as the interactive media to be used

by an elderly or handicapped person seated in an autonomous electronic wheelchair (see Fig. 4) that can carry its user to desired locations within complex environments autonomously.

The Architecture of MIGSEP.

The architecture of MIGSEP is illustrated in Fig 5. The Unified Dialogue Manager, which was developed according to the introduced unified dialogue model and the FormDia framework, functions as the central processing unit of the MIGSEP system and supports a flexible and context-sensible yet formally tractable multimodal interaction management. An Input Manager receives and interprets all incoming messages from the GUI Action Recognizer for GUI input events, the Speech Recognizer for natural language instructions and the Sensing Manager for other possible sensory data. An Output Manager on the other hand, handles all outgoing commands and distributes them to the View Presenter for presenting visual feedbacks, the Speech Synthesizer to generate and utter natural language responses and the Action Actuator to perform necessary motor actions, such as sending a driving command to the autonomous electronic wheelchair. The Knowledge Manager, which is closely connected with the unified Dialogue Manager, uses a Database to keep the static data of certain environments and the Context to process the dynamic information exchanged with the users during the interaction.

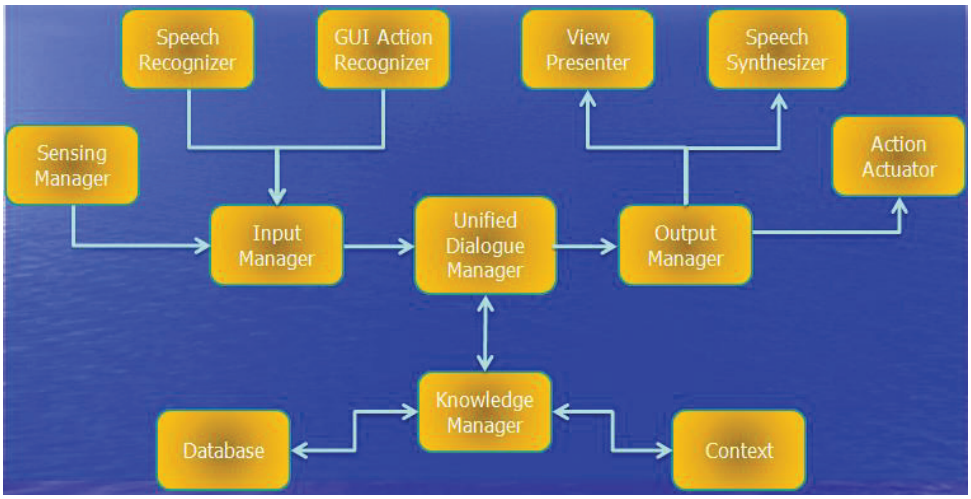


Fig. 5. The architecture of MIGSEP

The communication between the components of MIGSEP uses a uniform XML-protocol and each component can be treated as an open black box which can be accordingly modified or extended for specific use, without directly affecting other components in the MIGSEP architecture. It provides a general open platform for both theoretical research and empirical studies on single- or multimodal interaction that can relate to different application domains and scenarios

Multimodal Interaction with MIGSEP.

The current instance of MIGSEP has been set in the application domain of a simulated hospital environment. Fig.6 shows the configuration of the MIGSEP system where a user can use spoken-language, gesture via touch-screen or the combination of both to interact with MIGSEP.

Fig. 6. The current instance of MIGSEP

This system consists of a tablet PC, on which the MIGSEP instance is running and the interface is displayed, a button device for triggering a "press to talk" signal, and a green lamp to signalize the "being pressed and ready to talk" state, The MIGSEP interface itself can be divided into the following two areas:

- **Function-area** contains the function button "start" on the top left for going to the start state, the function button "toilet" showing the most basic need of an elderly person, and the text area for displaying the system responses;
- **Choice-area** displays information entities as single cards that can be clicked, with a scrollbar indicating the position of the current displayed cards and a context sensitive colored bar showing the current concerned context if necessary.

Fig. 7 shows an example of interaction between MIGSEP and a user who would like to go to the registration room of the endocrinology department.

Sys: <shows 3 cards, green for persons, yellow for rooms and blue for departments>
"Good day, I can help you find the desired location to go to."
User: "I want to go to the endocrinology departement."
Sys: <shows a blue bar texted with endocrinology on the top, and one brown card, one green card and one yellow card in the middle>
"Do you want me to show all the persons or the rooms in the endocrinology?"
User: <press the first card>
Sys: <enlarges the first brown card, shows yes/no button>
"Do you want to go to the registration room of endocrinology?"
User: <press the yes button>
Sys: <resizes the first brown card and then shows the 3 cards in the start menu again>
"OK, I have saved this goal. Where else do you want to go?"

Fig. 7. A sample interaction between a user and MIGSEP

3 Experimental Studies

In order to evaluate how well elderly persons can be assisted by MIGSEP system regarding different input modalities, i.e. gesture via touch-screen, speech, or the combination of both (abbreviated as combi), a series of experimental studies were conducted with the elderly persons in the fixed age range and the same experiment settings, which will be introduced in this section.

3.1 Participants

Altogether 31 elderly persons (m/f: 18/13, mean age of 70.7, standard deviation 3.1), all German native speakers, participated in the study. Every participant had to finish the mini-mental state examination (MMSE), a screening test to measure cognitive mental ability (cf. [36]). The participants are having the score of 29.0 averagely (std.=.84), which indicates that they have slight decline in cognitive abilities.

3.2 Stimuli and Apparatus

As shown in Fig. 6, visual stimuli were presented via a graphical user interface on the screen of a portable tablet PC and a green lamp, which is on if the speech input is activated; audio stimuli were generated by the MIGSEP and played via two loudspeakers at a well-perceivable volume. All tasks were given as keywords on the pages of a calendar-like system.

Three kinds of input possibilities were used to interact with MIGSEP: 1) speech input, activated if a button was being pressed and the green lamp was on, and deactivated if the button was released; 2) gesture via touch-screen, which is directly performable on the touch-screen display; 3) the combi modality, which allows participants to freely choose between speech and gesture via touch-screen as input modality.

The same data set contains virtual information about personnel, rooms and departments in a common hospital, was used in the experiment.

During the experiment each participant was accompanied by the same investigator, who gave an introduction to the system as well as predesigned instructions to interact with the system.

An automatic internal logger was implemented inside MIGSEP and used to record all the real-time system internal data, while an audio recorder program kept the whole audio data during the dialogic interaction process.

An evaluation questionnaire, focusing on the user satisfaction with MIGSEP regarding the subjective assessment using the combi modality, was especially designed for this study. It contains 6 questions concerning the quality of using the combi modality compared to the single modalities, each of which concerns with the feasibility, the advantages, the usability, the appropriateness and the preference. This questionnaire was answered by each participant via a five point Likert scale.

3.3 Procedure

At the beginning a brief introduction was given to the participants, so that they could get the basic idea and an overview of the experiment.

Then in order to minimize the learning or bias effect with respect to the use of one modality, we introduced a cross-over procedure with altogether three test runs, in which 16 participants out of 31 had to first use the gesture via touch-screen and then the speech input, and then the combi modality, while the other 15 first used speech and then the gesture via touch-screen, and finally the combi modality. Before each test run each participant was given comprehensive instructions and enough time to get familiar with each to be tested modality and its interaction with MIGSEP. During each test run each participant had to perform 11 tasks, each of which contained incomplete yet sufficient information about a destination the participant should select. For example a task can be to drive to "room 2603", to "Sonja Friedrich", or to "room 1206 or room 2206 with the name OCT-Diagnostics". For each modality the tasks were different at the content level, yet similar at the complexity level. Each task was fulfilled or ended, if the goal was selected or the participant gave up trying after six minutes.

After all tasks were run through, each participant was asked to fill in the questionnaire for evaluation.

4 Results and Discussion

According to the classic evaluation framework Paradise ([37]), the performance of an interactive system is determined by the effectiveness of task success, the efficiency of interaction and the user satisfaction about the system. Therefore, these three criteria are also used to analyze the data of the multimodal interaction between elderly persons and MIGSEP, while focusing on comparing the effects of using speech, gesture via touch-screen and combi input modalities.

4.1 Regarding Effectiveness of Task Performing

In the classic Paradise framework, kappa coefficient is used to measure the effectiveness of task performing of an interactive system. However, in order to calculate the kappa coefficient, a confusion matrix is needed and the original way of constructing a confusion matrix can only be applied to an interactive system with a single modality, which is not suitable for the data collected during multimodal interaction. Therefore, in order to construct the needed confusion matrix, we proposed the concept of an attribute value tree (AVT) in [14] to replace the attribute value matrix in the classic Paradise framework, where an AVT contains all information to be exchanged during the multimodal interaction between MIGSEP and elderly persons and therefore can function similarly to the original attribute value matrix, yet with the ability of dealing with multimodal interaction data.

As shown in Fig. 8, an AVT already contains all the information exchanged by using either gesture via touch-screen or speech input and thus, the 11 AVTs used in [14]

for assisting in evaluation the combi modality can also be used for evaluating the two single modalities.

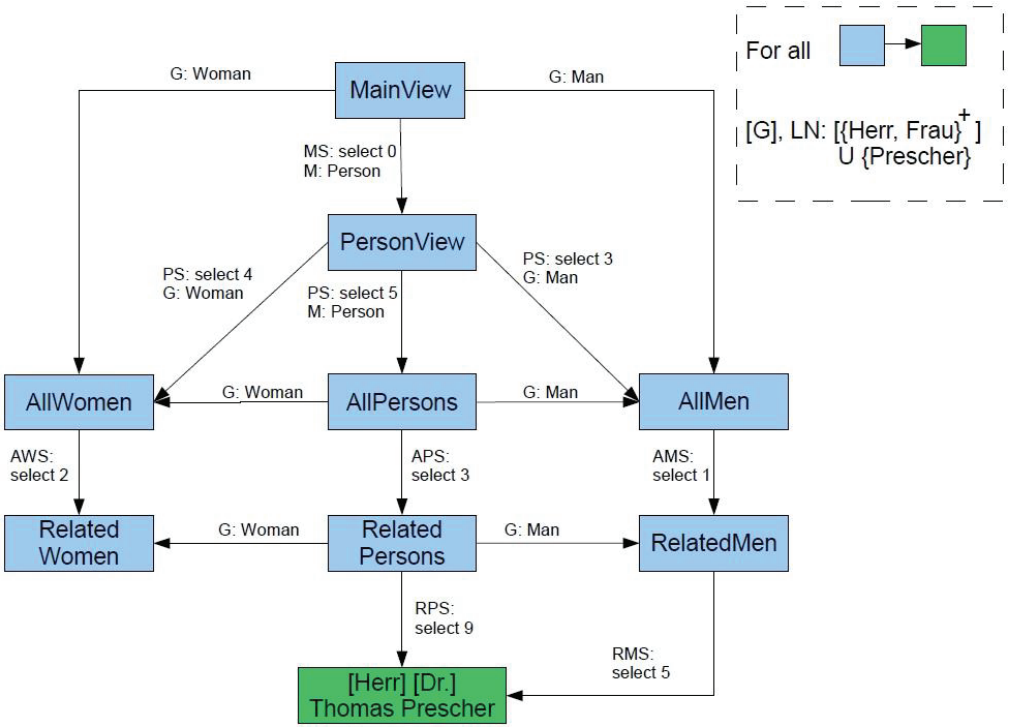


Fig. 8. An Attribute Value Tree for the task "go to Dr. Prescher"

Thus, the confusion matrix can then be constructed for all tasks and all modalities. For example table 1 shows a confusion matrix for the task "go to Dr. Prescher", where "M" and "N" denote whether the actual data match with the expected attribute values in the AVTs. There were 23 correctly selected actions in the MetaSelect (MS) state; the spoken language command regarding the last name (LN) was misrecognized by the system for 4 times; and the MetaType(M) "Person" was wrongly selected for one time. Note that, because of the width of the text, not all attributes of this confusion matrix can be shown in this example.

Table 1. The confusion matrix for the task "go to Dr. Prescher"

	MS		LN		…	M		sum
Data	M	N	M	N		M	N	
MS	23							23
LN			33	4				37
…								…
M						34	1	35

The data for all confusion matrices were merged and final confusion matrices of the 11 performed tasks for all the three modalities were created. Given the final confusion matrices, the Kappa coefficients were calculated with $\kappa = \frac{P(A)-P(E)}{1-P(E)}$ ([37]),

where $P(A) = \frac{\sum_{i=1}^{n} M(i,\ M)}{T}$ is the proportion of times that the actual data agree with the attribute values, and $P(E) = \sum_{i=1}^{n}(\frac{M(i)}{T})^2$ is the proportion of times that the actual data are expected to be agreed on by chance, where M(i, M) is the value of the matched cell of row i, M(i) the sum of the cells of row i, and T the sum of all cells.

Table 2. The kappa coefficients of task performing regarding the three modalities

	Speech	Touch-screen	Combi
κ	0.74	0.88	0.91
Std.	0.13	0.10	0.04

As shown in table 2, the three overall kappa coefficients show a sufficiently successful interaction between MIGSEP and the participants. The κ of speech input is 0.74, which is a bit lower than the gesture via touch-screen with κ=0.88, this is mainly caused by the automatic recognition errors. However, κ of Combi modality is even higher, indicating that with the combi modality the ASR errors were reduced considerably. The standard deviation 0.04 also implies that task performing using the combi modality is much more stable than the other two single modalities.

4.2 Regarding Efficiency of Interaction

Table 3 presents the results calculated from the quantitative data automatically recorded during the interaction using all the three input modalities, with respect to the most important factors for the efficiency analysis of an interactive system: the user and system turns in one dialogue for performing a task, the time used to complete a task and the number of speech recognition errors.

Table 3. Quantitative results regarding interaction efficiency for each participant and each task

	Touch-screen		Speech		Combi	
	Mean	Std.	Mean	Std.	Mean	Std.
User turns	15.5	4.1	4.3	1.7	7.4	3.6
Sys turns	15.4	3.9	4.3	1.6	7.4	3.6
Elapsed time (s)	88.9	40.2	57.6	24.2	48.7	20.0
ASR error times	-	-	0.8	0.7	0.3	0.4

With only half of the user/system turns and about 40 seconds less per participant per task, the interaction using combi modality shows a much better efficiency in all factors while comparing with gesture via touch-screen.

Although only averagely 4.3 user turns and 4.3 system turns were needed for each task using speech input, the error times of the automatic speech recognition (ASR) are much higher than using combi modality, which caused the 9 more seconds than using the combi modality for each participant while performing each task. This is also reflected in the lower standard deviation for combi modality, indicating that less extreme speech recognition problems occurred with combi modality than pure speech input.

4.3 Regarding User Satisfaction About The Modalities

The results of the questionnaire regarding the subjective comparison between the combi modality and the single modalities were analyzed and summarized in table 4.

Table 4. The subjective comparison between the combi modality and single modalities.

	Mean	Standard deviation
Better than single modality?	4.4	1.1
Easier solving tasks?	4.0	1.3
Showing advantages?	4.5	1.0
Usable to use combi-modality?	4.1	1.5
Prefer to use combi-modality?	4.4	1.3
Not confusing?	4.5	0.9
Overall	4.3	1.0

A very good overall user satisfaction with the combi modality is displayed via the score of 4.3 out of 5. Specifically, the elderly participants found the combi modality better and easier for performing tasks than the single modality with the score of 4.4 and 4.0; they could see the advantages of using combi modality and regard it as usable with the score of 4.5 and 4.1; they would prefer to use the combi modality with the score of 4.4; and they didn't find using the combi modality confusing with the score of 4.5. However, the scores of easier solving tasks and the usability of the combining modality were a bit lower than the others and the corresponding standard deviations were also a bit higher. Given a further insight into the data, this is mainly due to two elderly participants, who only used gesture via touch-screen even though both the modalities are enabled, and had made unpleasant impression of using only that, and therefore gave comparably lower score in the questionnaire.

5 Conclusion and Future work

This paper reported our work on multimodal interaction for elderly persons regarding the following two important aspects:

- The brief presentation of our theoretical and technical foundation for supporting designing, developing and implementing elderly-centered multimodal interactive systems;
- The summary and analysis of empirical data of a series of empirical studies for evaluating a practical multimodal interactive guidance system for elderly persons to be used in navigation scenarios in complex buildings, in which three modalities, speech, gesture via touch-screen and the combination of both, were compared.

In general, the overall positive results showed high effectiveness of task performing, high efficiency of interaction and high user satisfaction with the implemented system, which validated our systematically developed and empirically improved design guidelines, foundations, models and frameworks for supporting multimodal interaction for

elderly persons. The comparison between the input modalities also showed that the combination of gesture via touch-screen and speech input performs much better, more efficiently and is much more preferred by the elderly participants.

The presented work continued the pursuit of our final goal of building effective, efficient, adaptive and robust multimodal interactive systems and frameworks for elderly persons in Ambient Assisted Living environments. Corpus-based supervised and reinforcement learning techniques will be applied to support and improve the formal language driven unified dialogue modelling and management approach. Further application domain of autonomous navigation for elderly or handicapped persons on more spatially-related interaction is being investigated with qualitative spatial reasoning based frameworks. Special focus will also be placed on multimodal interaction within a smart home environment with a fully equipped technological infrastructure for the elderly or people with disabilities.

Acknowledgements. We gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG) through the Collaborative Research Center SFB/TR8, the department of Medical Psychology and Medical Sociology and the department of Neurology of the University Medical Center Göttingen.

References

1. Oviatt, S.: Multimodal Interfaces. In: Jacko, J.A., Sears, A. (eds.) The Human-Computer Interaction Handbook, pp. 286-304. L. Erlbaum Associates Inc., Hillsdale, NJ, USA (2002)
2. Jaimes, A., Sebe N.: Multimodal human-computer interaction: A survey. In: Computational Vision and Image Understanding, pp. 116-134. Elsevier Science Inc., New York, USA (2007)
3. Goldin-Meadow, S.: Hearing Gesture: How Our Hands Help Us Think. Harvard University Press (2005)
4. Kendon, A.: Gesture: Visible Action as Utterance. Cambridge University Press, Cambridge (2004)
5. Vitense, H.S., Jacko, J.A., Emery, V.K.: Multimodal feedback: establishing a performance baseline for improved access by individuals with visual impairments. In: Proceedings of the fifth international ACM conference on Assistive technologies (Assets '02), pp. 49-56. ACM New York, NY, USA (2002)
6. Reeves, L.M., Lai, J., Larson, J.A., Oviatt, S., Balaji, T. S., Buisine, S., Collings, P., Cohen, P., Kraal, B., Martin, J.-C., McTear, M., Raman, T.V., Stanney, K.M., Su, H., Wang, Q.Y.: Guidelines for Multimodal User Interface Design. In: Communication of the ACM – Multimodal Interfaces that flex, adapt, and persist, vol. 47(1), pp. 57-59. ACM New York, NY, USA (2004)
7. Lutz, W., Sanderson, W., Scherbov, S.: The coming acceleration of global population ageing. In: Nature 451, pp. 716-719. International Institute for Applied Systems Analysis, Laxenburg, Austria (2008)
8. D'Andrea, A., D'Ulizia, A., Ferri, F., Grifoni, P.: A multimodal pervasive framework for ambient assisted living. In: Proceedings of the 2nd International Conference on PErvasive

Technologies Related to Assistive Environments (PETRA '09), pp. 9-13. ACM, New York, NY, USA (2009)
9. Margetis, G., Antona, M., Ntoa, S., Stephanidis, C.: Towards Accessibility in Ambient Intelligence Environments. In: Proceedings of Ambient Intelligence - Third International Joint Conference (AmI 2012), pp. 328-337. Springer Berlin Heidelberg (2012)
10. Anastasiou, D., Jian, C., Zhekova, D.: Speech and gesture interaction in an Ambient assisted living lab. In: Proceedings of the 1st Workshop on Speech and Multimodal Interaction in Assistive Environments, pp. 18-27. Association for Computational Linguistics (ACL), Stroudsburg, PA, USA (2012)
11. Jian, C., Scharfmeister, F., Rachuy, C., Sasse, N., Shi, H., Schmidt, H., Steinbüchel-Rheinwll, N. v.: Towards Effective, Efficient and Elderly-friendly Multimodal Interaction. In: Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments, pp. 45:1-45:8. ACM, New York, USA (2011)
12. Jian, C., Scharfmeister, F., Rachuy, C., Sasse, N., Shi, H., Schmidt, H., Steinbüchel-Rheinwll, N. v.: Evaluating a Spoken Language Interface of a Multimodal Interactive Guidance System for Elderly Persons. In: Proceedings of the Fifth International Conference on Health Informatics, pp.87-96. SciTepress, Vilamoura, Algarve, Portugal (2012).
13. Jian, C., Shi, H., Scharfmeister, F., Rachuy, C., Sasse, N., Schmidt, H., Hoemberg, Steinbuechel, N.v.: Touch and Speech: Multimodal Interaction for Elderly Persons. In Biomedical Engineering Systems and Technologies. In: Lecture Notes, Communications in Computer and Information Science, pp. 385-400. Springer-Verlag, Berlin (2013).
14. Jian, C., Shi, H., Scharfmeister, F., Rachuy, C., Sasse, N., Schmidt, H., Steinbuechel, N.v.: Better Choice? Combing Speech and Touch in Multimodal Interaction for Elderly Persons. In: Proceedings of the 6th International Conference on Health Informatics. SciTepress, Barcelona, Spain (2013)
15. Fozard, J.L.: Vision and hearing in aging. In: Birren, J., Sloane, R., Cohen, G.D. (eds) Handbook of Metal Health and Aging, vol. 3, pp. 18-21. Academic Press (1990)
16. Mackay, D., Abrams, L.: Language, memory and aging. In Birren, J.E., Schaie, K.W. (eds), Handbook of the psychology of Aging, vol. 4, pp. 251-265. Academic Press (1996)
17. Moeller, S., Goedde, F., Wolters, M.: Corpus analysis of spoken smart-home interactions with older users. In: Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odjik, J., Piperidis, S., Tapias, D. (eds.) Proceedings of the Sixth International Conference on Language Resources Association, pp.735-740. ELRA (2008)
18. Kotary, L., Hoyer, W.J.: Age and the ability to inhibit distractor information in visual selective attention. In: Experimental Aging Research, vol. 21(2), pp. 159-171. Experimental Aging Research (1995)
19. McDowd, J.M., Craik, F: Effects of aging and task difficulty on divided attention performance. In: Journal of Experimental Psychology: Human Perception and Performance, vol. 14(2), pp. 267-280. American Psychological Association, Inc (1988)
20. Salthouse, T.A.: The aging of working memory. In: Neuropsychology, vol. 8(4), pp. 535-543. American Psychological Association, Inc (1994)
21. Craik, F., Jennings, J.: Human memory. In: Craik, F., Salthouse, T.A. (eds.) The Handbook of Aging and Cognition, pp. 51-110. Erlbaum (1992)
22. Hawthorn, D.: Possible implications of ageing for interface designer. In: Interacting with Computers, pp. 507-528 (2000)
23. Kline, D.W., Scialfa, C.T.: Sensory and Perceptual Functioning: basic research and human factors implications. In: Fisk, A.D., Rogers, W.A. (eds.) Handbook of Human Factors and the Older Adult, pp. 27-54. Academic Press (1996)

24. Schieber, F.: Aging and the senses. In: Birren, J.E., Sloane, R.B., Cohen, G.D. (eds.) Handbook of Mental Health and Aging, vol. 2. Academic Press, San Diego, CA (1992)
25. Charness, N., Bosman, E.: Human Factors and Design. In: Birren, J.E. Schaie, K.W. (eds.), Handbook of the Psychology of Aging, vol. 3, pp. 446-463. Academic Press (1990)
26. Smith, M.W., Sharit, J., Czaja, S.J.: Aging, motor control, and the performance of computer mouse tasks. In: Human Factors, vol. 41 (3), pp. 389-396 (1999)
27. Czaja, S.J., Sharit, J.: The influence of age and experience on the performance of a data entry task. In: Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting, pp. 144-147 (1997)
28. Sharit, J., Czaja, S.J., Nair, S., Lee, C.C.: Effects of age, speech rate, and environmental support in using telephone voice menu system. In: Human Factors, vol. 45, pp. 234-252 (2003)
29. Ziefle, M., Bay, S.: How older adults meet complexity: aging effects on the usability of different mobile phones. In: Behaviour and Information Technology, vol. 24(5), pp. 375-389 (2005)
30. Ross, R.J., Bateman, J., Shi, H.: Using Generalised Dialogue Models to Constrain Information State Based Dialogue Systems. In: the Symposium on Dialogue Modelling and Generation. Amsterdam, Netherlands (2005)
31. Traum, D., Larsson, S.: The information state approach to dialogue management. In: Kuppevelt, J.v., Smith, R. (eds.) Current and New Directions in Discourse and Dialogue, pp. 325-354. Kluwer (2003)
32. Shi, H., Bateman, J.: Developing human-robot dialogue management formally. In: Proceedings of Symposium on Dialogue Modelling and Generation. Amsterdam, Netherlands (2005)
33. Roscoe, A.W.: The Theory and Practice of Concurrency. Prentice Hall (1997)
34. Broadfoot, P., Roscoe, B.: Tutorial on FDR and Its Applications. In: Havelund, K., Penix, J., Visser, W. (eds.) SPIN model checking and software verification, LNCS vol. 1885, pp. 322. Springer-Verlag, London, UK (2000)
35. Mandel, C., Huebner, K., Vierhuff, T.: Towards an Autonomous Wheelchair: Cognitive Aspects in Service Robotics. In: Proceedings of Towards Autonomous Robotic Systems (TAROS2005), pp. 165-172 (2005)
36. Folstein, M., Folstein, S., Mchugh, P.: "mini-mental state", a practical method for grading the cognitive state of patients for clinician. In: Journal of Psychiatric Research, vol. 12(3), pp. 189-198 (1975)
37. Walker, M.A., Litman, D.J., Kamm, C.A., Kamm, A.A., Abella, A.: Paradise: a framework for evaluating spoken dialogue agents. In: Proceedings of the eighth conference on European chapter of Association for computational Linguistics, pp. 271-280. , NJ, USA (1997)

A Conceptual Model for Human-Robot Collaborative Spatial Navigation

Cui Jian and Hui Shi

SFB/TR8 Spatial Cognition
University of Bremen
Enrique-Schmidt-Str. 5, 28359, Bremen, Germany

Email: {ken, shi}@informatik.uni-bremen.de

Abstract

This paper describes our work on developing effective, efficient and user-friendly interaction between a human operator and a mobile robot on performing spatial navigation tasks. In order to solve the spatially related communication problems caused by the disparity between human mental representation about spatial environments and that of a mobile robot, a qualitative spatial knowledge based four-level conceptual model is proposed. With a computational framework based on an application dependent instance of this model, high-level conceptual strategies are implemented and used to support the human-robot collaborative spatial navigation. An empirical study is then conducted to evaluate the computational framework implemented into a practical interactive system using a real environment map regarding different conceptual strategies.

Keywords: Conceptual Modelling, Qualitative Spatial Representation and Reasoning, Communication of Spatial Information, Human-Robot Interaction.

1 Introduction

As intelligent service robots are receiving more and more attention in academic and industrial areas, considerable research efforts have been dedicated to the development of effective, efficient and user-friendly human-robot interaction in different application domains (Fong et al (2003), Goodrich and Schultz (2007)). The major concern of our work is placed on solving communication problems during human-robot interaction in the domain of spatial navigation, where a mobile service robot is collaboratively controlled by an intelligent embedded system for low-level autonomous navigation and a human operator for giving high-level conceptual route instructions using natural language. The human operator can tell the robot, for example, to turn around, go straight ahead, take a right, and then pass a coffee machine on the left, until it reaches the copy room.

Much research has been devoted in this area, e.g., (Koulouri and Lauria 2009) and (Shi and Tenbrink 2009) performed corpus-based analysis on natural language route directions with mobile robots; (Kollar, et al 2010) and (Marge and Rudnicky 2010) studied the relationship between features of spatial environment and language, especially the role of natural language in route instructions; (Zender, et al 2008) and (Mozos 2010) proposed and improved a multi-layered conceptual model corresponding to spatial and functional properties of typical indoor environments based on topological information, then used this model to support a mobile robot's indoor navigation. Diverging from these methods and models concentrating on empirical data, natural language and topological conceptual information, our work is focusing on human perspectives according to the following two important aspects.

First, in human-robot collaborative navigation, human operators usually use natural language expressions containing qualitative relations and conceptual landmarks (Hirtle 2008), such as "go to the end of the corridor, turn right, and then go until the coffee machine on the left", while mobile robots work on quantitative level and can only interpret instructions with quantitative data, such as "145.0 meters ahead, then make a 37.5 degree turning, …". There is apparently a gap between a human operator and a mobile robot if they want to communicate with each other. Much research has been focusing on applying mathematical well-founded qualitative spatial calculi and models to represent and reason about spatial environments (e.g. Ligozat and Renz (2004), Schultz, et al (2006), Wolter and Lee (2010), Kurfess, et al (2011)). Adding to this body of literature, using qualitative spatial knowledge as an intermediate layer for the intuitive human-robot communication has been viewed as the foundation of our work.

Furthermore, providing a sequence of route instructions is a rather complex process for the human operator, since spatially-related communication problems could easily occur if spatial objects are wrongly localized or a certain instruction is wrongly given due to a certain spatial situation, e.g., a coffee machine cannot be found after taking a right turn, or a room to be passed is not on the left as expected (Reason (1990) and Bugmann (2004)). Therefore, an effective mechanism is needed for the mobile robot and the human operator to collaboratively identify the problems and negotiate possible solutions with each other.

Thus, in order to bridge the interaction gap between the human operator's qualitative spatial mental model and the mobile robot's quantitative representation, as well as supporting the high-level collaborative negotiation of spatially-related communication problems, we proposed a

qualitative spatial knowledge based four-level conceptual model: the Qualitative Spatial Beliefs Model (QSBM). This model was first proposed in (Shi and Krieg-Brückner 2008), and then extended and implemented with a computational framework (Jian, et al 2009) and a set of high-level conceptual strategies to support collaborative human-robot spatial navigation (Shi, et al 2010). Two conceptual strategies were evaluated and compared in (Jian, et al 2010). With the further development of the conceptual model based computational framework and the integration into a practical interactive system for a mobile robot, the current paper reports on a new high-level conceptual strategy for resolving more spatially-related human-robot communication problems, as well as an empirical study, which was conducted to test the current system with the focus of evaluating the new conceptual strategy and its comparison with the previous strategies.

The remainder of the paper is organized as follows. Section 2 presents the qualitative spatial knowledge based four-level conceptual model and one of its application dependent instance with conceptual strategies to solve the spatially related communication problems. Section 3 introduces the computational framework that implements the conceptual model. Section 4 then describes the empirical study to evaluate a practical interactive system regarding the model-based conceptual strategies and Section 5 discusses the results of the study. Finally, Section 6 concludes the paper and gives an outline to our future work.

2 A Qualitative Spatial Knowledge based Four-Level Conceptual Model

2.1 The Overview of the General Model

According to the perspective of human operators, spatial environments are not represented with quantitative data as a mobile robot does, but with conceptual objects or places and their qualitative spatial relations. Accordingly, Qualitative Spatial Beliefs Model (QSBM), a qualitative spatial knowledge based conceptual model is developed to model a mobile robot's beliefs for supporting more intuitive communication with human operators. Figure 1 illustrates the general QSBM with a four-level structure.

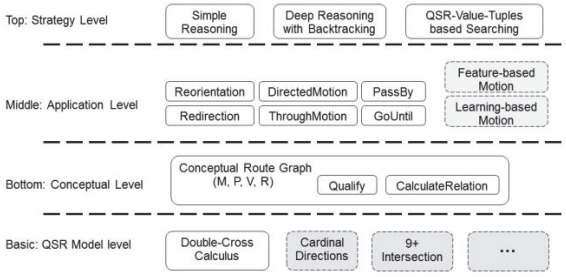


Figure 1: The QSR-based four-level conceptual model: Qualitative Spatial Beliefs Model (QSBM)

The basic level is the QSR Model level, which contains the most basic theoretical foundation of the QSBM model: qualitative spatial calculi for different application requirements, such as Double-Cross Calculus (Freksa, 1992), Cardinal Directions (Frank, 1991), 9+ Intersection (Kurata, 2008), etc.

Based on the chosen qualitative spatial calculus, a basic conceptual model can be constructed and serves as the fundamental conceptual level. This level only contains qualitative spatial information and the basic calculating and reasoning mechanism with respect to the connection between the chosen calculus and the navigation environment. It can be seen as a black box holding a conceptual qualitative spatial knowledge based representation of a spatial environment with two basic functions: *Qualify* for qualifying quantitative information into qualitative relations, and *CalculateRelation* for calculating additional qualitative spatial relations with qualitative spatial relations between objects using calculus-based qualitative spatial reasoning.

The application level consists of a set of most atomic application-dependent update rules, which correspond to all the possible user-uttered route instructions to a mobile robot in collaborative spatial navigation. For instance, the update rule *Reorientation* can refer to the instruction "turn left", *Redirection* can interpret "take the next junction on the left". *Feature-based Motion* concerns instructions with features of objects or landmarks, such as "go around the big laboratory" (see (Gondorf and Jian, 2011)), and *Learning-based Motion* represents those instructions requiring the robot to augment its conceptual knowledge by learning new landmarks or disambiguating landmarks, such as "the third office is the directory's office, pass by it", etc. Each update rule is used to update the state of the spatial representation on the conceptual level with respect to its formal definition based on a chosen calculus and the related qualitative spatial reasoning on the QSR Model level.

On the strategy level, high-level conceptual strategies are developed to assist in interpreting a sequence of route instructions and if possible, resolve the spatially-related communication problems during the collaborative spatial navigation. Basically, each conceptual strategy defines its own mechanism for appropriately choosing and applying atomic update rules on the application level.

In general, the QSBM is a conceptual model for applying qualitative spatial knowledge to represent a spatial environment, qualitative spatial reasoning to define a set of application-dependent update rules to update the conceptual representation, and conceptual strategies to manage the atomic update rules to support high-level spatially-related human-robot communication. With the flexibility and expandability provided by the multi-level structure, further application scenarios can be supported by using different qualitative calculi on the QSR model level, more application-dependent actions can also be added on the application level, or new high-level strategies can also be implemented to resolve more communication problems, while each of these changes/extensions requires only limited adaptation on the other levels in QSBM.

Specifically, since qualitative spatial calculi at the QSR Model Level are well studied, formal details about the other three levels of an instance of QSBM will be given in the rest of the chapter.

2.2 A DCC-based QSBM

Considering the current requirement of the collaborative spatial navigation scenarios, double-cross calculus (DCC) is selected as the basic QSR model and a DCC-based

QSBM is developed (Shi, et al 2010) and introduced according to the conceptual, application and strategy level as follows.

2.2.1 The Conceptual Level

In mobile robot navigation, one of the most important basic models is called Route Graph (Werner, et al 2000). Route graphs are a special class of graphs, with graph nodes representing conceptual places at geographical positions regarding a quantitative reference system, and graph edges or route segments, each of which is directed from one node to another and altogether build up a conceptual network of routes (see Fig. 2 a)). Conventional route graphs cannot only be used as quantitative representation of spatial environments for mobile robots' navigation, they also capture the topological knowledge of space from human perspective and therefore have the potential of intermediate layers for human operators. However, the gap between quantitative representation of conventional route graphs and qualitative knowledge based mental construct of human operators remains a problem preventing a more direct interaction.

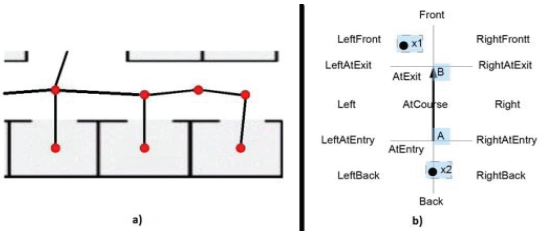


Figure 2: a) one part of a conventional route graph; b) the orientation frame of Double Cross Calculus with 15 qualitative spatial relations.

On the other hand, Double Cross Calculus (DCC (Freksa, 1992)) divides the 2-dimensional space with a directed segment into disjoint grids (see Fig. 2 b)), which defines 15 meaningful qualitative spatial relations. Thus, a DCC model can be used as a local navigation map from an egocentric perspective and support the interaction with human operators in a local navigation scenario.

By combining the structure of a conventional route graph and the DCC model, the conceptual route graph (CRG) is developed (Shi and Krieg-Brückner, 2008). A CRG inherits the topological structure from a conventional route graph, where quantitative information is completely replaced by the DCC relations between graph nodes and route segments. Formally, a CRG of a spatial environment is defined by a tuple of four elements (M, P, V, R):

- M is a set of landmark-place-pairs in the environment, specifying the locations of all the landmarks at places in P, such as an {office: x_1}, or a {kitchen: x_2}.
- P is a set of topological places, or the graph node in a CRG, such as x_1 or x_2.
- V is a set of vectors, each of which is directed from one place to another place, such as *AB*.
- R is a set of relation-pairs, which specify the DCC relations between places and vectors. A relation pair is written as <AB, LeftFront, x_1>, meaning that x_1 is in the LeftFront grid of AB.

Therefore, the CRG for the simple spatial environment illustrated in Fig. 2 b) is represented as:

```
crg = (M = {office:x1,kitchen:x2},
       P = {A, B, x1, x2},
       V = {AB, BA},
       R = {<AB, LeftFront, x1>, <AB, Back, x2>})
```

And a state of a DCC-based QSBM model, which is stored as a mobile robot's internal representation about current spatial situation, can then be represented for example as:

```
<crg, pos = AB>
```

This means that the mobile robot is now located at place A and looking at the direction of place B, with an office on the LeftFront position and kitchen at the Back.

2.2.2 The Application Level

In order to support the application scenarios of human-robot collaborative spatial navigation, a set of route instructions such as "turn left", "take the next junction on the right", "pass by the office on the left", etc., should be interpreted by the mobile robot. According to the formal definition of the DCC-based CRG on the conceptual level, a set of low-level update rules regarding the most common route instructions for mobile robots are developed on the application level and used to update the state of the DCC-based QSBM, i.e., the state of a mobile robot about spatial environment.

Each update rule is specified with the following three elements:

- a name (followed by RULE), which identifies a class of most common route instructions,
- a set of preconditions (followed by PRE), under which this update rule can be applied, and
- an effect (followed by EFF), describing how the state of the DCC-based QSBM is updated after applying the update rule.

As examples, the update rules for reorientation and directed motion are presented as follows:

- **Reorientation** refers to the simplest route instructions, which change the current orientation of a robot, such as "Turn left", "Turn right" and "Turn around". In general, the precondition is whether a robot can find a CRG vector satisfying two conditions: 1. it is originated from the current place and 2. It is targeted at a place that has the desired spatial relation with the current position; the effect is that the robot position is updated as that found CRG vector, formally described as:

```
RULE: Reorientation
PRE: pos = P0P1,
     ∃P0P2∈V. <P0P1, dir, P2>
EFF: pos = P0P2
```

Concretely, the rules indicates that the robot is currently at the place P_0 and faces the place P_1 (P_0P_1 is a CRG vector), if there exists a CRG vector P_0P_2 with a targeting place P_2, such that the spatial relation of P_2 with respect to the route

segment P_0P_1 (i.e. the current position) is the desired direction dir to turn, i.e., $<P_0P_1$, dir, $P_2>$, then the current position will be updated as P_0P_2 after applying this update rule.

- **Directed Motion** defines the class of the route instructions that usually contain a motion action and a turning action changing the direction of the continuing motion, such as "take the next junction on the right". These instructions usually involve with a landmark (e.g. the "junction"), until which the robot should go, and a direction (e.g. on the "right"), towards which the robot should turn. For example, in general, for the route instruction "take the next corridor on the right", the first corridor on the right from the robot's current position needs to be identified first. Thus, the update rule for directed motions with the first landmark and a turning direction is specified as:

RULE: DirectedMotionWithFstLandmarkAndDir
PRE: pos = P_0P_1,
$\exists P_2P_3 \in V$. $((l:P_2) \wedge <P_1P_2, dir, P_3> \wedge <P_0P_1, Front, P_2>)$
$\wedge \forall P_4P_5 \in V$. $((l:P_4) \wedge <P_1P_4, dir, P_5> \wedge <P_0P_1, Front, P_4>$
$\wedge (P_2 \neq P_4)) \rightarrow <P_1P_2, Front, P_4>$
EFF: pos = P_2P_3

In this rule, l is the targeted landmark and dir is the direction to turn to; The first precondition specifies that the robot should find a CRG vector P_2P_3, such that the targeted landmark is located at P_2, the spatial relation between P_3 and the segment P_1P_2 is the desired direction dir and P_2 is in front of the robot's current position; The second precondition limits that P_2 is the first place referring to the given landmark at the given direction, instead of an arbitrary one; this condition is satisfied if there exists a place P_4 with the same feature as P_2, P_4 must be ahead of P_2 from the current perspective. The effect is that, the robot position is updated to P_2P_3 after applying this rule. Similarly, other variants of directed motions, such as "go straight ahead", "go right" or "take the second left" can be specified with similar update rules accordingly.

2.2.3 The Strategy level

With the update rules defined on the application level, single route instructions can be interpreted. However, in human robot collaborative navigation, human operators usually give a sequence of route instructions to the mobile robot. In this case, if a certain route instruction is wrongly given, spatially related communication problems could easily occur, because taking the wrong route instruction could cause problems of interpretation of the subsequent route instructions, which could result in failure of the entire interpretation or even lead to a completely unexpected route.

In order to resolve these problems, a set of high-level conceptual strategies are developed on the strategy level, which apply the low-level update rules accordingly and appropriately according to different principles and methods. Among them, the two most important conceptual strategies are briefly introduced as follows.

2.2.4 Reasoning with Backtracking

With the qualitative spatial reasoning on the QSR model level, the preconditions of update rules on the application level can easily be checked, this is in fact the most straightforward way to see if a sequence of route instructions can be interpreted. However, there are often situations where the failure of the interpretation of some instructions is caused by a previously incorrect instruction, e.g. see the situation in Fig. 3. The robot is located at the thick red arrow and the instructions are: "go straight ahead, then go left, and then turn right, and go until the kitchen on the right." A simple check fails on interpreting the fourth instruction "go until kitchen on the right", because there is no kitchen ahead after taking a right turn as the previous instruction. However, by taking one step backwards, if the third instruction is changed from right to left, then the last instruction can also be interpreted properly.

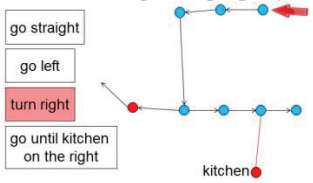


Figure 3: An example of a wrong instruction

Thus, the strategy "Reasoning with Backtracking" (abbr. RwB) interprets the route instructions as the straightforward way does, checking every precondition as usual. Yet after applying each update rule, the state of the updated QSBM is also saved in an interpretation history. Once one instruction cannot be interpreted, the previous state of the QSBM can be reloaded as the current state and possible suggestion can be made based on the previous instruction, such as "turn left" instead of "turn right" in the example in Fig. 3. As a result, the checking of the preconditions of the remaining route instructions can be resumed based on the suggested route instruction, and a possible route matching the entire sequence of route instructions can be found.

The RwB strategy has been evaluated and compared with other conceptual strategies and the positive results were reported in (Jian, et al 2010).

2.2.5 QSR-Value Tuples based Searching

During the development and integration of the QSBM model into an interactive system to be used by a mobile robot, a new class of spatially-related communication problems is identified. Fig. 4 illustrates one example of these problems.

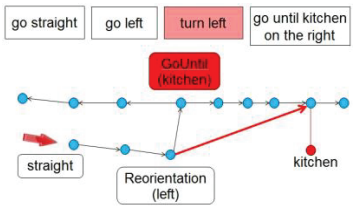


Figure 4: An example of a "missing" instruction

In this example, the robot is located at the thick red arrow and the instructions are "go straight, then left, then go until the kitchen on the right". From the perspective of the human operator, the kitchen is located directly on the right side, and therefore the operator simply ignores a turning point that is in the conceptual representation but not in his/her mental representation. However, after taking a right turn, the last instruction "go until kitchen on the right" cannot be interpreted, because there is no continuing possibility as shown in Fig. 4.

These problems cannot be solved by the RwB strategy, because the RwB strategy can only provide suggestions if there exists a wrong route instruction, while in these situations one route instruction is missing. Thus, the strategy "QSR-Value Tuples based Searching" (abbr. QSRVT) was developed. For each outgoing direction of each turning node in a conceptual route graph during the interpretation, a QSR weighted value tuple is defined as:

```
(route, instructions, qsr_v)
```

where *route* is the currently taken route, *instructions* is the set of all the along this route interpreted instructions, and *qsr_v* is the cumulative value calculated by

$$qsr\_v = \sum_{i=0}^{i_{current}} mr_i * sr_i$$

where mr_i is the matching rate by comparing the desired qualitative spatial direction with the current route direction while interpreting the *i*-th instruction, sr_i is the success rate of interpreting the *i*-th route instruction, and $i_{current}$ is the index of the current route instruction.

The QSRVT strategy first initializes an empty set of QSR-value tuples at the starting position of the robot. This set of QSR-value tuples is then automatically updated and expanded by the searching agents of the QSRVT strategy, while they are travelling along all paths (according to the branching of each turning node) in the QSBM. Finally, a full set of QSR-value tuples is generated and the QSR-value tuple with the highest QSR-weighted value is either the best possible solution for interpreting the route instructions or contains the most relevant information to provide possible suggestion to resolve the spatially-related communication problems.

As an example, Fig. 5 briefly illustrates how the QSRVT strategy solves the problem in Fig. 4.

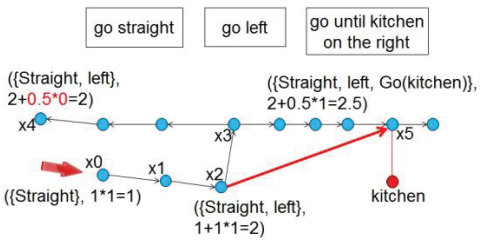


Figure 5: A simple process of the QSRVT strategy

After interpreting the first two instructions "go straight" and "go left", the searching comes to the turning node x3. There are two possible directions going out of x3 and accordingly two more QSR-value tuples are added. In this situation, the last instruction cannot be interpreted with the left going route while it can be interpreted with the right going one. Therefore, the instructions are interpreted with the route x0->x1->x2->x3->x5, since the QSR-value tuple has the highest value 2.5.

3 A Conceptual Model based Computational Framework

Based on the introduced QSBM, including update rules and the high-level conceptual strategies, we developed SimSpace, a conceptual model based computational framework for supporting the implementation of QSBM into a practical interactive system to be used by a mobile robot.

3.1 General Architecture

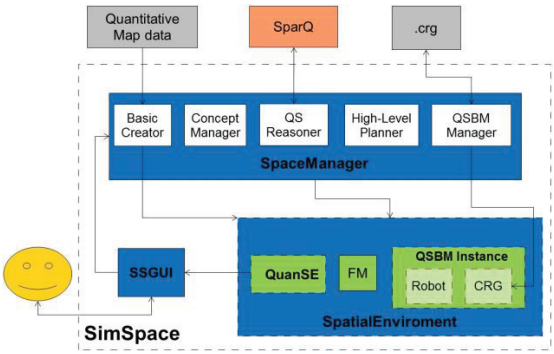


Figure 6: The general architecture of SimSpace

According to the Model-View-Controller architecture (originally from (Burbeck 1987)), the general architecture of SimSpace consists of a Model component *Spatial Envioronment*, an optional View Component *SSGUI* and a Controller *SpaceManager*:

Spatial Environment maintains the current state of the QSBM instance, i.e., the conceptual route graph and the hypothesis of the robot position in the CRG, as well as the optional quantitative spatial environment (QuanSE) for quantitative data and the optional feature map (FM) component containing the conceptual information.

SSGUI is the graphical user interface of SimSpace. It is an optional component and is only used if the SimSpace system is started as a stand-alone application. It visualizes the spatial environment with quantitative and conceptual descriptions, interacts with a human user who is giving the natural language route instructions, and communicates with the Space Manager for the interpretation of incoming route instructions as well as outgoing system responses.

Space Manager is the central processing component of SimSpace, it consists of the following five functional components:

- Basic Creator creates a spatial environment instance with quantitative and conceptual data according to the quantitative map data, if given.
- Concept Manager manages an ontology database of the conceptual knowledge, such as names of landmarks or persons, how they are conceptually related, etc. It is used to interpret the conceptual terms in the natural language route instructions.

- QS Reasoner is connected with SparQ (Wolter and Wallgrün 2011), a general toolbox for qualitative spatial representation and reasoning. It supports the most basic operations on the conceptual level in QSBM, e.g., qualification of quantitative data into qualitative relations and calculation of qualitative spatial relations.
- QSBM Manager connects with QS Reasoner and generates a QSBM instance according to a qualitative spatial calculus on the QSR model level and a quantitative environment if given, manipulates and updates an empty or existing QSBM instance with the application dependent update rules on the application level, and saves the updated QSBM instance into a XML-based specification with .crg file extension, if needed.
- High-Level Planner implements the high-level conceptual strategies to apply appropriate update rules to interpret route instructions and resolve spatially-related communication problems.

3.2 The Interpretation of Route Instructions in SimSpace

The SimSpace system can interpret a sequence of human route instructions in the following steps:

- The sequence of route instructions is firstly parsed into a list of predefined semantic representations.
- According to the activated high-level conceptual strategy, each semantic representation is assigned with an applicable low-level update rule.

  For each low-level update rule, its preconditions are instantiated. Taking the sample instruction "go until the kitchen on the right" in the previous section, the update rule GoUntilRight is applied and by substituting the current robot position with the CRG vector *AB* and the location of the kitchen is found as P_{kit}, the second precondition is instantiated to:

$$\exists P_2P_3 \in V.\ (\text{kitchen}:P_{kit}) \wedge \langle AB, \text{RightFront}, P_{kit}\rangle \wedge \langle P_2P_3, \text{RightBack}, P_{kit}\rangle \wedge \langle AB, \text{Front}, P_2\rangle \wedge \langle AB, \text{Front}, P_3\rangle$$

  Then with the support of the SparQ toolkit, the instantiated preconditions are checked against the current state of the QSBM.

  If the current state matches the instantiated precondition, the current robot position is updated to P_2P_3 and a message object containing the success information is returned.

  If the current state provides e.g. the relations:

$$(\text{kitchen}:P_{kit}) \wedge \langle AB, \text{LeftFront}, P_{kit}\rangle$$

  This means, the kitchen is located on the left side from the perspective of the robot and therefore, $\langle AB, RightFront, P_{kit}\rangle$ in the precondition cannot be satisfied. In this case, SimSpace creates a corresponding message which contains necessary information for indicating the failure of the interpretation and/or generating suggestion.
- According to the conceptual strategy and the returned message, either the interpretation continues if possible, or strategy dependent process is performed (e.g. in the RwB or QSRVT strategy), or appropriate responses or suggestions are made and presented back to the human user.

On the one hand, SimSpace can be used as a stand-alone evaluation platform for visualizing spatial environments, generating corresponding QSBM instances and testing the interpretation of natural language route instructions. On the other hand, it can also be used as a well encapsulated module and integrated into an interactive system to be used by a mobile robot to assist in the interaction with human operators.

4 An Empirical Study

In order to evaluate the qualitative knowledge based conceptual model and its implementation into a practical interactive system regarding the two different high-level conceptual strategies: reasoning with backtracking and QSR-value-tuples based searching, an empirical study was conducted.

4.1 Participants

Altogether 18 university students, with no background knowledge on cognitive science and therefore considered as novice users, participated in the study, in which 9 of them were interacting with the system using the strategy reasoning with backtracking, while the other 9 were testing the system with QSR-value-tuples based searching.

4.2 Stimuli and Apparatus

All stimuli were the same for each participant during the interaction process, e.g., visual stimuli were presented on a graphical user interface on a laptop displaying a map of an indoor environment with named landmarks, a robot avatar showing the current position of the robot, a possibly highlighted route along which the robot is going, and the clearly emphasized text of system response with respect to participants' instructions (see Fig. 7); audio stimuli of the system response were also generated as complementary feedback and played via the external speaker of the same laptop at a well-perceivable volume.

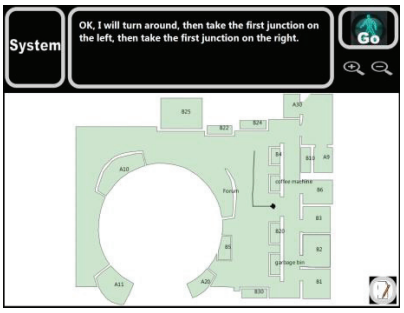


Figure 7: The graphical user interface with all visual stimuli

The same map of a floor plan of an indoor environment with the same virtual landmarks within this environment was used throughout the study.

The interactive system was a networked software system consisting of two laptops: one laptop, called the system laptop, hold the actual interactive system, which included the graphical user interface, interaction manager, speech synthesizer and the spatial knowledge processing component SimSpace that implemented the qualitative knowledge based conceptual model and the conceptual strategies; the other laptop, called the speech recognizer laptop, run a graphical interface, which was only operated by a human investigator and used to transfer the spoken natural language instructions to the system laptop via wireless network. The time for inputting natural language instructions is significantly shortened with a well-designed group of function-buttons on the speech recognizer laptop, so that only two seconds on average were needed for transferring utterances to the system laptop. As a result, the whole system was simulated as if each participant was giving instructions to the system using spoken natural language directly.

All participants were accompanied by the same investigator, who gave the introduction to the study and the system at the beginning, and input the natural language instructions of each participant into the speech recognizer laptop during the task performing through pressing the function buttons.

An internal automatic logging program of the system was used to collect interaction data such as dialogue turns, utterances, event time, and so on, while the standard audio recorder of windows recorded the whole dialogic interaction process.

Two questionnaires were conducted. The first one is called spatial ability questionnaire, which includes questions regarding abilities of describing routes to others, inquiring ways from others and using map in everyday life. This questionnaire aims to get the subjective assessment of each participant about his/her cognitive spatial abilities; the second questionnaire is called evaluation questionnaire, which concerns with the user satisfaction with the interactive system. Both questionnaires were based on 5-point Likert scale.

4.3 Procedure

For each test a participant had to undergo four steps:

1. Self-assessment: the participant was asked to fill the spatial ability questionnaire.
2. Introduction: the participant was given a brief introduction to the system and the following test runs, which included how to interact with the system and what to expect during the interaction.
3. Interaction: each participant was given five different tasks, each of which contains a starting position and a goal position. Only spoken language instructions were used to tell the robot to go from the starting position to the goal position. In order to collect more data and to produce more problem situations, for each task the participant had to describe two different routes or utter two different descriptions. Each task was ended, if either the goal position was reached, or the participant gave up trying.
4. Evaluation: after interacting with the system, the participant was asked to fill in the evaluation questionnaire.

5 Results and Discussion

According to the general view of the well accepted evaluation framework Paradise (Walker, et al 1997), the performance of an interactive system can be measured via the effectiveness, the efficiency and the user satisfaction. Thus, we have performed the analysis of the data from the interactive system on the two conceptual strategies with respect to these three aspects.

Even with the relatively small group of the participants (9 persons in each group), the authors believed that the comparison of the presented empirical results between the two groups can be considered representative, since the grouping was performed in a random manner, and furthermore, the results of the self-assessment of the spatial ability are similar between the two groups with the values of 53.2 and 51.9 on average.

5.1 Regarding the Effectiveness

The study was conducted with a Wizard of Oz setting without an automatic speech recognizer, therefore, the effectiveness of the interactive system could only depend on whether the subtasks were successfully performed, namely, whether the navigation goals were reached or not. 10 Goals were supposed to be reached by each participant. With 9 participants for one strategy, the number of reached goals are counted and summarized in table 1.

	RwB	QSRVT
Reached Goals (percentage)	85 (94.4%)	90 (100%)

Table 1: Effectiveness with RwB and QSRVT

For both strategies, the effectiveness of performing navigation tasks with the interactive system is very good. The participants using the RwB strategy reached 85 goals out of 90, while the ones using the QSRVT strategy reached all the goals.

5.2 Regarding the Efficiency

In order to find out how efficiently each participant was assisted with the interactive system using the two different strategies, the automatically logged data were analysed according to the average elapsed time and interaction turns for each task. The results are summarized in table 2.

	RwB		QSRVT		P Value
	Mean	Std.	Mean	Std.	
Average Elapsed Time (s)	87.37	33.13	48.12	9.14	0.007
Average Interaction Turns	7.14	2.91	4.07	0.68	0.013

Table 2: Data concerning efficiency for each participant and each task

From a general perspective for task performing, a very good efficiency is shown with 87.37 seconds and 7.14

interaction turns on average for each task with the RwB strategy, since this also includes some very long system responses, some of which even needed over 20 seconds to be played. The standard deviation of 33.13 for elapsed time is however a bit high, this is mainly due to one certain participant who confused the left/right relations too often and used over 150 seconds on average to finish one task, which, however, is not common for the other participants.

Moreover, the performance efficiency with the QSRVT strategy is much better: each participant only used 48.12 seconds and 4.07 turns on average to finish one task. The p-values of 0.007 and 0.013 also indicate that the participants with the QSRVT strategy could perform tasks significantly more efficiently than those with the RwB strategy.

5.3 Regarding the User Satisfaction

Regarding the user satisfaction about the interactive system, the subjective data of the evaluation questionnaire filled by each participant after task performing were analysed and summarized in table 3.

	RwB		QSRVT	
	Mean	Std.	Mean	Std.
System Response	3.36	0.79	4.0	0.52
General Support	3.94	0.80	4.25	0.41
Future use	3.72	0.49	3.94	0.62
Total	3.68	0.63	4.06	0.43
Total / Skill	0.07	0.02	0.08	0.01

Table 3: Data concerning user satisfaction

The overall user satisfaction of the interactive system with the RwB strategy for each participant is considered at a satisfactory level with the total average value of 3.68 and standard deviation 0.63. Specifically, they found the system response sufficiently understandable with the value 3.36, they felt supported by the system with the value of 3.94 and they would recommend the system with the value of 3.72. The standard deviations of 0.79 for system response and 0.80 for general support are a bit higher, this is because of the special situations where the RwB strategy encounters with missing instructions and therefore the system could not provide very useful information about the communication problems.

Meanwhile, the user satisfaction of the system with the QSRVT strategy was improved from every perspective, 4.0 for the system response, 4.25 for the general support, 3.94 for the future use and all together 4.06.

With the data from the self-assessment questionnaire, a skill value is calculated and shows how confident each participant considers him- or herself to be with spatially-related tasks. The ratios of the total satisfaction degree and the skill value of 0.07 and 0.08 also roughly indicate that, the QSRVT strategy better assists the participants also in a more or less subjective manner than the RwB strategy does.

6 Conclusion and Future Work

In this paper we reported our work on using conceptual model to support human robot collaborative navigation, focusing on the following three important aspects:

- the design and development of a qualitative spatial knowledge based multi-level conceptual model for human robot interaction,
- the implementation of the conceptual model and the model-based high-level conceptual strategies within a general computational framework, and
- the evaluation of an interactive system built on the conceptual model, framework and strategies.

The positive empirical results validated our effort on developing and implementing the proposed conceptual model and framework. It was also shown that, the model based high-level conceptual strategies, especially the strategy of QSR-value tuple based searching can assist the mobile robot to clarify more spatially-related communication problems and better support the human-robot collaborative navigation.

The presented work served as a fundamental step towards building robust, effective, efficient, user-friendly models, frameworks and interactive systems in spatially-related applications. The integration of the conceptual model, framework and strategies into a real mobile robot for spatial navigation with untrained human operators is being conducted. For the strategy of QSR-value tuple based searching, learning-based QSR-value updating is being investigated. We are also considering adding other qualitative spatial calculi into the QSR model level to support further application, such as object localization within complex buildings. Human-robot collaborative exploration in unknown or partially-known spatial environments is also a work package to be pursued.

Acknowledgement

We gratefully acknowledge the support of the German research foundation (Deutsche Forschungsgemeinschaft, DFG) through the Collaborative Research Center SFB/TR8 spatial cognition and the University of Bremen.

7 References

Bugmann, G., Klein, E., Lauria, S., Kyriacou, T. (2004): Corpus-based robotics: A route instruction example. *Proc. Of the 8th International Conference on Intelligent Autonomous Systems*, Amsterdam, Netherlands, 96-103.

Burbeck, S. (1987): Applications Programming in Smalltalk-80 (TM): How to use Model-View-Controller (MVC).http://st-www.cs.uiuc.edu/users/smarch/st-docs /mvc.html.

Fong, T., Nourbakhsh, I., and Dautenhahn, K.(2003): A survey of socially interactive robots. *Robotics and Autonomous Systems* **42**(3-4): 143-166.

Frank, A. U. (1991): Qualitative Spatial Reasoning with Cardinal Directions. Proceedings of the seventh Austrian Conference on Artificial Intelligence, 157-167, Springer Berlin Heidelberg.

Freksa, C.(1992): Using Orientation Information for Qualitative Spatial Reasoning. *Proceedings of the International Conference GIS – From Space to Territory: Theories and Methods of Spatio-Temporal Reasoning on Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, 162-178, Springer-Verlag.

Gondorf, C.P.J. and Jian, C. (2011): Supporting Inferences in space – a Wayfinding task in a multilevel building in space – a Wayfinding task in a multilevel building. *Proc. of the 2nd Workshop on Computational models of Spatial Language Interpretation and Generation*. 48-51. Hois, J., Ross, R.J., Kelleher, J.D., and Bateman, J.A. (eds).

Goodrich, M.A. and Schultz, A.C. (2007): Human-robot interaction: a survey. *Foundations of Trends in Human-Computer Interaction* **1**(3): 203-275.

Hirtle, S.C. (2008): Landmarks for navigation in human and robots. *Robotics and Cognitive Approaches to Spatial Mapping, Springer Tracts in Advanced Robotics*, 38: 203-214, Springer Berlin Heidelberg.

Jian., C., Shi, H., Krieg-Brückner, B. (2009): A Tool to Interpret Route Instructions with Qualitative Spatial Knowledge. In *AAAI Spring Symposium: Benchmarking of Qualitative Spatial and Temporal Reasoning Systems*. 47-48.

Jian, C., Zhekova, D., Shi, H., Bateman, J. (2010): Deep Reasoning in Clarification Dialogues with Mobile Robots. *Proc. of the 2010 Conference on 19th European Conference on Artificial Intelligence (ECAI)*. 177-182. Coelho, H., Studer, R. and Wooldridge, M. (eds). IOS Press.

Kollar, T., Tellex, S., Roy, D., and Roy, N. (2010): Toward understanding natural language directions. *Proc. of the 5th ACM/IEEE international conference on Human-robot interaction*, NJ, USA, 259-266, IEEE Press.

Koulouri, T. and Lauria, S. (2009): A corpus-based analysis of route instructions in human-robot interaction. In *Towards Autonomous Robotic Systems (TAROS)*, Londonderry, 281-288, University of Ulster.

Kurata, Y. (2008): The 9+ Intersection: A Universal Framework for Modelling Topological Relations. In: Geographic Information Science, Lecture Notes in Computer Science 5266: 181-198, Springer Berlin Heidelberg.

Kurfess, F., Flanagan, G., and Bhatt, M. (2011): Spatial Interactions between Humans and Assistive Agents. *Proc. of the AAAI 2011 Spring Symposium: Help Me Help You: Bridging the Gaps in Human-Agent Collaboration*, 42-47, AAAI.

Ligozat, G. and Renz, J. (2004): What is a qualitative calculus? A general framework. In: *Trends in Artificial Intelligence-8th Pacific Rim International Conference on Artificial Intelligence*. 53-64. Zhang, C., Guesgen, H.W. and Yeap, W.K. (eds). Springer.

Marge, M. and Rudnicky, A.I. (2010): Comparing Spoken Language Route Instructions for Robots across Environment Representations. *Proc. of SIGDIAL 2010: the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, Tokyo, Japan, 157-164, Association for Computational Linguistics.

Mozos, O.M. (2010): Conceptual Spatial Representation of Indoor Environments. *Semantic Labeling of Places with Mobile Robots*, *Springer Tracts in Advanced Robotics*, 61: 83-97, Springer Berlin Heidelberg.

Reason, J. (1990): Human Error. Cambridge University Press.

Schultz, C.P.L., Guesgen, H.W., and Amor, R. (2006): Computer-human interaction issues when integrating qualitative spatial reasoning into geographic information systems. *Proc. of the 7th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: design centered HCI*, New York, NY, USA, 43-51, ACM.

Shi, H., Jian, C., and Krieg-Brückner, B. (2010): Qualitative Spatial Modelling of Human Route Instructions to Mobile Robots. *Proc. of the Third International Conference on Advances in Computer Human Interactions*, Washington DC, USA, 1-6, IEEE Computer Society.

Shi, H. and Krieg-Brückner, B. (2008): Modelling Route Instructions for Robust Human-Robot Interaction on Navigation Tasks. *International Journal of Sofware and Informatics* 2(1): 33-60.

Shi, H. and Tenbrink, T. (2009): Telling Rolland Where to Go: HRI Dialogues on Route Navigation. In *Spatial Language and Dialogue*. 177-190. Conventry, K., Tenbrink, T. and Bateman, J. (eds). Cambridge University Press.

Walker, M.A., Litman, D.J., Kamm, C.A., Kamm, A.A., and Abella, A. (1997): PARADISE: A Framework for Evaluating Spoken Dialogue Agents. *Proc. of the eighth conference on European Chapter of Association for Computational Linguistics*, NJ, USA, 271-280.

Werner, S., Krieg-Brückner, B., and Herrmann, T. (2000): Modelling Navigational Knowledge by Route Graphs. In: *Spatial Cognition II, Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications*. 295-316. Freksa, C., Brauer, W., Habel, C. and Wender, K.F. (eds). Springer-Verlag, London, UK, UK.

Wolter, D. and Wallgrün, J.O. (2011): Qualitative Spatial Reasoning for Applications: New Challenges and the SparQ Toolbox (final draft version), In: *Qualitative Spatio-Temporal Representation and Reasoning: Trends and Future Directions*. Hazarika, S. M. (eds).

Wolter, D. and Lee, J.H. (2010): Qualitative reasoning with directional relations. *Artificial Intelligence* 174(18): 1498-1507.

Zender, H., Mozos, O.M., Jensfelt, P., Kruijff, G.-J.M., Burgard, W. (2008): Conceptual spatial representations for indoor mobile robots. *International Journal of Robotics and Autonomous Systems* 56(6): 493-502, Amsterdam, Netherlands.