TRITA-CSC-A 2009:02
ISSN-1653-5723
ISRN–KTH/CSC/A--09/02-SE
ISBN 978-91-7415-224-1

ANDERS GREEN Designing and Evaluating Human-Robot Communication

KTH 2009

www.kth.se

Designing and Evaluating Human-Robot Communication

Informing Design through Analysis of User Interaction

ANDERS GREEN

Doctoral Thesis in
Human-Computer Interaction
Stockholm, Sweden 2009

KTH
VETENSKAP
OCH KONST

KTH Computer Science
and Communication

to Samuel and Edvin for showing me what actually matters...

Abstract

This thesis explores the design and evaluation of human-robot communication for service robots that use natural language to interact with people. The research is centred around three themes: design of human-robot communication; evaluation of miscommunication in human-robot communication; and the analysis of spatial influence as empiric phenomenon and design element.

The method has been to put users in situations of future use through means of Hi-fi simulation. Several scenarios were enacted using the Wizard-of-Oz technique: a robot intended for fetch- and carry services in an office environment; and a robot acting in what can be characterised as a home tour, where the user teaches objects and locations to the robot. Using these scenarios a corpus of human-robot communication was developed and analysed.

The analysis of the communicative behaviours led to the following observations: the users communicate with the robot in order to solve a main task goal. In order to fulfil this goal they overtake service actions that the robot is incapable of. Once users have understood that the robot is capable of performing actions, they explore its capabilities.

During the interactions the users continuously monitor the behaviour of the robot, attempting to elicit feedback or to draw its perceptual attention to the users' communicative behaviour. Information related to the communicative status of the robot seems to have a fundamental impact on the quality of interaction. Large portions of the miscommunication that occurs in the analysed scenarios can be attributed to ill-timed, lacking or irrelevant feedback from the robot.

The analysis of the corpus data also showed that the users' spatial behaviour seemed to be influenced by the robot's communicative behaviour, embodiment and positioning. This means that we in robot design can consider the use strategies for *spatial prompting* to influence the users' spatial behaviour.

The understanding of the importance of continuously providing information of the communicative status of the robot to it's users leaves us with an intriguing design challenge for the future: When designing communication for a service robot we need to design communication for the robot work tasks; and simultaneously, provide information based on the systems communicative status to continuously make users aware of the robots communicative capability.

Acknowledgements

A PhD thesis is intended to describe an individual effort. In this respect, a thesis concerning Human-Robot Interaction is really a non sequitur. It can never happen without the collaboration between humans! I want to mention and thank a whole bunch in a particular, but largely insignificant, order.

My parents, who have never stopped believing in me and have supported me in just about whatever I have tried to do: things like riding tri- and bicycles, flyfishing or diving in ponds, getting in the way from footballs, hockey-pucks, handballs, etc, or embarking on something really strange, like convincing other people that I am trying to study speaking robots.

Helge Hüttenrauch my companion into the uncharted territory of Human-Robot Interaction. Elin Anna Topp, who brought things further by actually making robots do things for real. Erik Espmark, who turned the rather whimsical idea of a robot doll into true living art. Lars Oestreicher who provided many good ideas during the Cero projekt. Mikael Norman, who made invaluable efforts in the Cero project. Patric Jensfelt, who explained peculiar things about robots so that even I understood. Britta Wrede, Manja Lohse, Shu-yin Li, Marc Hanheide, Mick Walters and Nuno Otero for making collaboration in the Cogniron project enjoyable and fun. Fredrik Olsson for allowing me to use his photo on the front cover. Anette Arling, Jeanna Ayobi, Ulla-Britt Lindqvist and Karin Molin for heroic administrative efforts. And of course the people at the HCI group at KTH and the fellow students and staff in the Gradaute school of Human-Machine Interaction and the Graduate School of Natural Language Technology.

My supervisors Kerstin Severinson Eklundh and Henrik Christensen who both have the unique capacity of questioning perfectly self-explainable ideas and thereby forcing me to turn them into something which share similarities to research.

And last but not least, Maria Cheadle, for providing unfathomable emotional, intellectual and contextual support. Yes, its time to say “Finally!”.

Contents

CHAPTER 1

Introduction

This thesis is about service robots that use natural language to interact with people. The underlying assumption for this work is that *human-to-human communicative behaviour* can be used as a basis, or inspiration, for the design of interaction for service robots. In the following I will refer to speaking robots as having an *interaction model* based on human natural language. The interest in natural language as an interface model comes from the assumption that a robot which is to be operated by ordinary people in everyday environments requires an interaction model that is intuitive, efficient and reliable. The basic assumption for this is that a service robot which offers an interaction model that matches human language performance in terms of conveying and understanding complex meaning will be perceived as intuitive, efficient and satisfactory by its users, at least to the same extent that interaction with people can be said to have these characteristics.

Human communicative behaviour provides a highly complex and rich web of different behaviours and characteristics which provide research challenges that are interesting in their own right. To some extent this has led to a scientific paradigm which promotes research with a narrow focus, concentrated on models and methods for handling specific phenomena related to human natural language. Based on the expectations of research on natural language processing, interfaces that emulate human communicative behaviour have been advocated as a means of giving direct and intuitive support for the user's actions. Karsenty (2002) has noted that this narrow focus on human-like behaviour and capabilities has stimulated research

on natural language interfaces that focuses on achieving systems with "perfect performance". This is similar to the situation in research on humanoid robotics and socially interactive robots, where research often is focused on models and methods for imitating and emulating specific aspects of human behaviour that contribute to the *appearance* of the robot.

The view taken in this work is that when the interactive capability provided by software components that emulate and imitate human behaviour becomes part of the task repertoire of a service robot it is necessary to incorporate development and evaluation efforts that address both task performance and communicative capability[1]. More specifically, the goal for this thesis is to investigate how an interaction model for service robots, based on human communicative behaviour, can be designed and evaluated in a realistic use context.

1.1 A multidisciplinary research process

The challenge of providing user interfaces for service robots can be a approached from different perspectives. The design of a communicative interface requires an understanding of human-robot communication as well as of techniques for developing multimodal natural language interfaces. In my view this cannot be done in one step. Instead design of *user interfaces* for robots is seen as a *multidisciplinary process* where design and research ventures benefit from each other.

Another important focus for the research presented in this thesis is to consider the perspective of the user during the development process. Understanding the needs, motivations and concerns of users are key challenges for human-robot communication, and it has been a persistent goal to involve them at every possible stage in the design process.

From a more technical point of view, a *user* can be seen as an agent attempting to achieve certain goals using the robot as a sophisticated tool. The notions of *task* and *use* then become important: the user *uses* the robot in order to solve a specific *task* or use a service provided by the system. A task is understood as something that the robot primarily performs using its *physical* capabilities, even if it is possi-

[1] The distinction between task performance and communicative capability is not straightforward from a philosophical point of view. Service task are actions and if we adhere to the notion of Austin (1962) that *language is action*, we need to treat communication just as any other service offered by a robot.

ble for the robot as a language user to perform actions through verbal means. In the following I will use the term *participant* to denote persons that are invited to interact with robots as users in our studies. When I refer to human-robot interaction design in general terms or when describing system actions and behaviour from a system perspective I will use the term *user*. As this thesis is concerned with aspects of use rather than social or psychological aspects of human-robot interaction I have refrained from using the term *human*, unless human qualities are being specifically referred to. Here I adhere to what appears to be a well-established terminology, used for instance in the extensive survey by Fong, Nourbakhsh and Dautenhahn (2003a) on social robots.

Research context

In practical terms the work described in this thesis has been carried out during the years 1999 – 2008, in the context of two projects. The first project, started in 1998, concerned the development of an office robot, Cero, initiated as a project together with the Swedish National Labour Market Board (AMS), but mainly financed by the Swedish Foundation for Strategic Research (SSF), the Swedish Graduate School of Language Technology, and *Swedish Transport and Communications Research Board* (KFB)[2].

The second project, "The Cognitive Robot Companion (Cogniron)", financed by the European Commission, started in 2004 and ended in 2008. The Cogniron project was focused on research methods for sensing, moving and acting, focusing on the development of cognitive and social capabilities necessary for a type of robot that was characterised as a "cognitive companion". The capabilities of such a robot include focusing of attention, understanding of the spatial and dynamic structure of the environment, together with communicative functions that allow it incorporate and appropriate social behaviour in a given context (Cogniron, 2003).

The interest for our group has been focused on a robot demonstrator, a *Key Experiment* that was to show central capabilities of the robot companion. Using this key experiment as a basic scenario we have explored research challenges concerning ways of interactively providing information to a robot companion through a so called *Home Tour*. In the home tour scenario a user and robot interact to de-

[2]Now VINNOVA (Swedish Governmental Agency for Innovation Systems)

fine objects and locations in the user's home. The objectives of the key experiment provided a rich research context in which the ideas described in this thesis could be explored.

My research in the Cogniron project was carried in two interconnecting research activities concerning *multi-modal dialogues* and *social behaviour and embodied interaction* which were carried out in close cooperation with the University of Bielefeld (Germany) and the University of Hertfordshire (UK).

When I started, around 1999, research on human-robot interaction with service robots was a relatively new and marginal field of academic research. Very few (if any) service robots were commercially available on the consumer market and there was only small a number of research platforms available. Today the number of available research platforms has grown and there are now several types of robots available from a large number of companies. The field of Human-Robot Interaction research has also grown. Until a few years ago the IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man) was one of few conferences that focused on human-robot interaction. Now even the major robotics conferences such as the IEEE International Conference on Robotics and Automation (ICRA) and IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) naturally include papers with technical aspects of Human-Robot Interaction. The human-computer interaction community endorsed by the ACM has also become interested in human-robot interaction through workshops in the ACM CHI conference (CHI2004). In 2006 ACM launched the annual ACM Human-Robot Interaction Conference (HRI), which manifests Human-Robot Interaction as a research discipline.

The next level of maturity of human-robot interaction research can be seen at the horizon with the careful launching of a few commercially available service robots, allowing for new types of studies (Forlizzi, 2007) on a growing mass consumer market (Jones, 2006).

1.2 Research approach

In the following I will approach human-robot communication from two perspectives:

- The first concerns *human-robot communication design to support users' interaction with semi-autonomous service robots*.
- The second perspective focuses on *how qualitative analysis and evaluation of use in realistic scenarios can inform design of human-robot communication*.

Design of task-oriented dialogue for service robots

The first research challenge addressed in this thesis concerns the investigation of human-robot communication design for task-oriented, autonomous service robots. The overall goal is to establish what properties and qualities of communicative behaviour of humans are required of robots in order to achieve a level of usability that allows practical use. The method I have chosen to approach this is to design, build and analyse communicative interfaces for task-oriented service robots. The focus on the design-oriented aspects of this research should be seen in the light of the fact that there are still very few commercially available robot systems that include a user interface that supports natural language dialogue. Precursors to robots with natural language user interfaces are instead found in laboratories (for instance, Breazeal et al. 2005; Haasch et al. 2004), museums and science fairs (for instance, Schulte et al. 1999; Siegwart et al. 2003).

When humans communicate, they engage in joint communicative behaviour with the purpose to establish and maintain common ground (Clark, 1996). One main assumption for this work is that human-robot communication has many characteristics in common with task-oriented human-human dialogue. This is partly because humans are involved, but also because the robot uses natural language as a vehicle for exchanging and sharing information about joint goals.

During the initial stages of the development process for the office robot Cero (see Green 2001; Green et al. 2000; Green and Severinson Eklundh 2003 and Hüttenrauch et al. 2004), we realised that there was more to human-robot communication than verbal dialogue concerning specification of tasks to be solved by the robot:

- The communication between humans and robots is multimodal, incorporating verbal utterances, gestures, gaze, positioning and posture.
- The embodiment of the robot, its appearance and its movements influence the behaviour and attitudes of the user.
- The environment in which the robot acts, the shared space between the user, the location and the objects available forms a complex use scenario.
- The communicative feedback given by the robot influences the quality of the interaction.

Both in human-human communication and in human-computer interaction providing feedback is important for the interactive process. In human-human communication feedback provides the means for participants in conversation to jointly acquire common ground, for instance by providing evaluations of contributions by means of displayed multimodal communicative behaviour (Allwood, 2002; Allwood et al., 1991). The creation of common ground also involves the manner in which dialogue participants configure the shared context. The body and the immediate environment are used as an interactive locus for the creation of meaning and action (Goodwin, 2000).

In human-computer interaction it is generally assumed that feedback during interaction is essential for the usability of a system. By receiving informative feedback from the system the user is becoming aware of system states and actions that are performed by the system. Appropriate feedback in user interfaces reduces disorientation and confusion of users (Shneiderman and Plaisant, 2004).

Another challenge for the design of human-robot communication is evaluation. First of all we need to find ways to establish quality criteria for human-robot communication as such. This can be achieved through analysis intended to inform design but also by providing the means to detect and repair miscommunication.

Secondly, we need to establish quality criteria for how robots can achieve a level of communication that allows them to provide useful services.

The challenges listed above can be summarised in these research questions that I will try to answer in this thesis:

- What is the appropriate communication design for an autonomous service robot? What are the relevant practical and theoretical aspects?
- How can we analyse and evaluate the quality of human-robot communication?

Corpus-based evaluation in the design process

The second research challenge concerns how to analyse situated human-robot communication with respect to communicative quality and communicative functions. For this purpose I am using a corpus-based approach to support the development process of natural language user interfaces. In the course of the work we have employed the Wizard-of-Oz technique to collect data on how users act and behave when faced with a personal service robot. The resulting corpus not only contains data on verbal and gestured communication but also spatial configurations and information on tasks. Taken together a corpus of this kind provides a rich context for analysis of human-robot communication as it represents interaction which is unfolding in several concurrent tracks allowing for studies of multimodal interaction. The research in this thesis has utilised the corpus for two main areas: *to analyse and categorise miscommunication* to inform design and to understand how the robot can influence the spatial behaviour of the robot, exploring the concept of *spatial prompting*.

In the corpus data I have observed sequences of interaction that display symptoms of miscommunication, defined as a state of misalignment between the mental states of agents involved in communication. This means that either the speaker fails to produce the effect intended with the communicative acts issued or the hearer fails to perceive what the speaker intended to communicate (Traum, 1996).

Even though some parts of this thesis concern design of practical dialogue systems, on-line detection and repair[3] of miscommunication has not been in focus when designing these system. The miscommunication analysis described in this

[3]For an excellent overview of these aspects see (Skantze, 2007)

thesis is largely qualitative, and performed as an integral part of the design process (see Chapter 6). The primary goal is to improve the system as it is being redesigned in an iterative development process. The research concerning corpus-based analysis of human-robot communication and the subsequent analysis of miscommunication has been motivated by the following research questions:

- How can corpora of human-robot communication be used in the design and evaluation of human-robot communication? How can we categorise and analyse communicative behaviours?
- What are the types and characteristics of miscommunication in human-robot communication? How can we design human-robot communication to reduce or prevent miscommunication?

Influencing spatial behaviour of users

The third research focus related to observations made in the corpus, concerns the observation that the robot *was actively influencing the users' spatial behaviour*. This led to a discussion that ended in the use and conceptualisation of the term *spatial prompting* (see Green and Hüttenrauch 2006). While most accounts of spatial adaptation in robot systems are focused on the robot's adaptation to the human movement, the interest in this thesis concerns how the robot actively can influence the spatial behaviour of the user. Spatial prompting can be used to create a spatial configuration between the user and the robot that is beneficial for the purposes of the ongoing interaction. An example would be to suggest a position that would facilitate detection of gestures or spoken input through deliberate communicative behaviour and movements by the robot (this is further exemplified in Chapter 8).

Situated communication between a mobile service robot and its users takes place in a physically shared environment, and typically concerns entities and activities that can be referenced, viewed and manipulated by the participants. In human-to-human contexts, behaviour that seeks to actively influencing the spatial positioning of one another is used as a natural ingredient of social interaction and can range from unreflected actions such as occupying space and thereby making others change their position to deliberately pushing or tackling someone. Some sports, like Ice hockey or American football provide good examples of the latter. People are mostly aware of the consequences of their spatial behaviour, for instance, they

know when they are in someones way. The assumption of this research is that in order to influence the spatial behaviour of others, robots needs to be explicitly designed to take their own spatial behaviour into consideration.

This thesis is concentrated on some aspects of how the robot actively can influence spatial behaviour of the user. The understanding of space has been studied in depth, for instance in social anthropology, and the term "spatial prompt" has been used in relation to the discussion of space syntax (Hillier and Hanson, 1984) and territoriality (Sack, 1986). Widlock et al (1999) uses the term "spatial prompt" to describe how a specific feature of a building, an *olupale*[4], projects change in social behaviour. In the following the term is used to capture phenomena that are related to actions that the robot can take to influence the behaviour of people. Phenomena related to spatial influence have been studied by Lewin (1939) who discussed the notion of *social forces*. Lewin's account of spatial influence has been used to model and simulate how pedestrians coordinate conflicts of spaces, like when passing a door opening and how they form lanes (Helbing and Molnár, 1995). The questions regarding spatiality I am focusing on in this thesis are:

- Can spatial prompting be motivated empirically?
- In what way can we design communicative behaviour of robots to influence the spatial behaviour of users?

1.3 What this thesis is not about

The goal of this thesis is to investigate human-robot communication with task-oriented service robots from a user centered design perspective. Creating a robot with a real task, and an interface with a robustness that would allow iterative development of user specific adaptations and long-term user studies in a full scale scenario is not within the scope of this thesis. The technical limitations of the robots and interface components that were available at the time of this research has limited the research to design studies, ranging from conceptualisation to simulated or rudimentary interface prototypes with limited capability. Moreover the work has nevertheless been focused on practical use of robots rather than more psychologically oriented research focused on the study of attitudes towards the appearance, behaviour and character of robots; and for example, the role of robot and

[4] which can be described as a fire place for guests, found in some villages in northern Namibia.

human personality for approach distances, anthropomorphisation and communicative behaviour which are all interesting phenomena with respect to human-robot interaction, and has been studied in a number of works (cf. Fussell et al. 2008; Syrdal et al. 2006; Walters et al. 2008).

1.4 Definitions of service robots

As this work concerns communicative service robots it is initially useful to discuss possible definitions of the term *service robot*. The International Federation of Robotics (IFR) has proposed this definition of a service robot: *"A robot which operates semi or fully autonomously to perform services that are useful to the well being of humans or equipment"* [5]. A problem with the IFR definition is that it does not provide a further definition of neither the term "service" nor, in fact, "robot". I therefore assume that a *service* is *work done for the benefit of another* or an *act of help or assistance*[6].

Another assumption is that a robot in this context is *a reprogrammable multifunctional mobile device* following the definitions[7] of ISO (8373) and the Robotic Industries Association (RIA) which both includes "reprogrammable" and "multifunctional" in their definitions. A possible problem with this definition is that it does not exclude simple systems with sensors and an actuators, like sliding doors or escalators that are activated when you step in front of them. In fact a lot of machines can be said to provide automatic services for users, ranging from coffee makers, dishwashers to door openers and automatic defibrillators. I am hesitant to call these machines *robots*. Instead the terms that describe these machine seem to derive directly from the service they provide, or they are affixed by words[8] like 'automatic', 'electronic', 'motorised' or 'mechanical' etc. At the heart of the matter lies that service robots display *autonomous behaviour* and that they may be used for *general purposes*, meaning that they can be programmed and re-programmed for different tasks. We could argue that a robot should be *multifunctional* to qualify as a service robot, but this would exclude single function robotic devices, like

[5] Definition from www.ifr.org (last checked: 2008-10-22)

[6] Both meanings are listed in http://wordnet.princeton.edu/

[7] Definition is quoted from in Encyclopædia Britannica. Retrieved October 21, 2008, from Encyclopædia Britannica Online: http://www.britannica.com/EBchecked/topic/44912/automation

[8] Ten synonyms (including 'robotic') are listed in Roget's New Millennium Thesaurus, First Edition (v 1.3.1). Lexicon Publishing Group, LLC. Accessed on 22 Apr. 2008.

vacuum cleaners and lawn mowers that seem to fit the description of robots in other respects, like being mobile and solving tasks autonomously. It seems that at the outer edges of conceptualisation we are left with intuition to determine what characterises a robot.

The robots in the scenarios I have worked with in this thesis are intended to be mobile, autonomous, reprogrammable and provide services for humans. The view taken in this work is also that "mobile" concerns movement in general, like the capability of the robot of transporting itself autonomously between different places. I also assume the perspective that a robot is able to *manipulate* its environment to some degree, even without having a specific device like a robotic arm attached. Manipulation is understood in this work in a very broad sense: by acting in an environment a robot can manipulate it by positioning itself in a certain place to influence the actions of other agents, by pushing things using its body, or by performing social acts through communication, for instance through the use of *speech acts* (Austin, 1962).

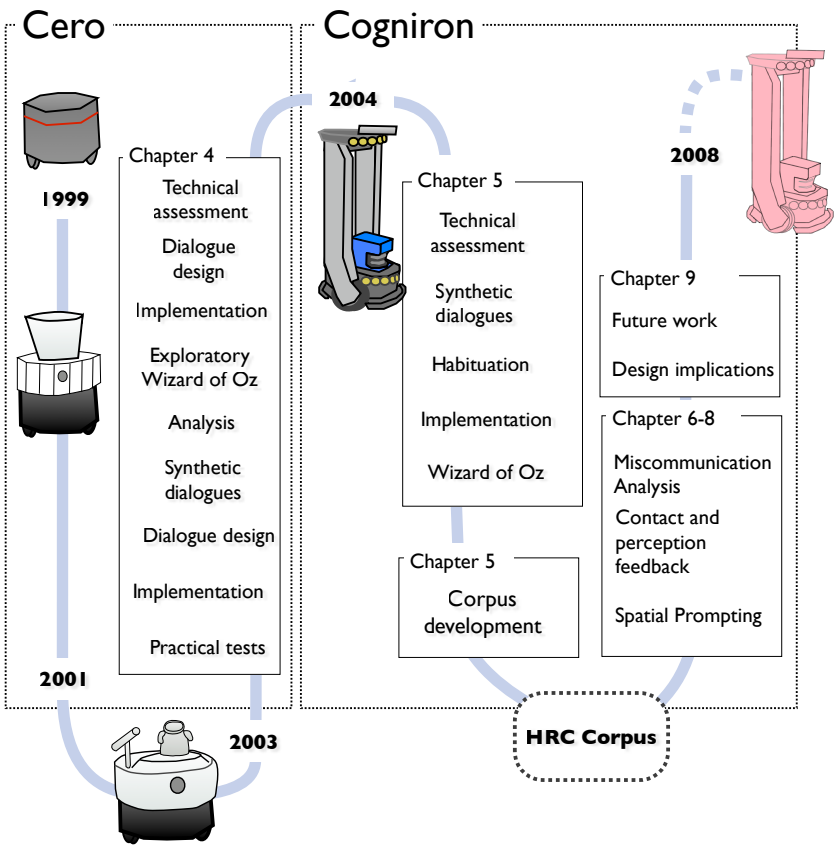


Figure 1.1 *Time-line of research themes, methods and outcome of activities.*

1.5 Thesis outline

The nine chapters of the thesis can be grouped into four parts. Initially I introduce human-robot communication as a research subject and provide an overview of different models for interaction and how this has studied and manifested in user interface design. Then I turn to the design, implementation and evaluation of the Cero robot system. The following chapters are focused on evaluation of human-robot communication in the European project Cogniron. In the last chapter the

result of this work is summarised and discussed. The following outline[9] sketches the purpose of each of the chapters:

Chapter 1, introduces the research focus and research questions.

Chapter 2 introduces interaction models for human-natural language dialogue and discusses their relation to human-robot communication design.

Chapter 3 gives a background to the methods for designing communication, elicitation of user behaviour and approaches for corpus based analysis used in the thesis.

Chapter 4 describes how the human-robot communication for the Cero project was designed and evaluated and what we learned from this process.

Chapter 5 concerns the communication design for an interactive scenario, the *Home Tour*, investigated in the Cogniron project. The chapter is focused on how the design was adapted to suit a real use situation and how this was used to elicit interactive behaviour to create a corpus of human-robot communication.

Chapter 6–7 addresses the research questions regarding communication design for an autonomous service robots and how this can be evaluated. The research question regarding how to collect and use corpora in the evaluation is approached both in chapter five and six. Chapter 6 also addresses the research questions regarding the types and characteristics of miscommunication.

Chapter 7 describes an analysis of contact and perception feedback in the corpus material and discusses implications for design of human-robot communication. This chapter is relevant for the research question regarding how we can increase the quality of communication to prevent miscommunication by dialogue design

Chapter 8 discusses the notion of spatiality in human-robot interaction and introduces and motivates the concept *spatial prompting* as a design element. This chapter addresses research questions regarding how we can understand spatiality in human-robot communication and more specifically how we can influence users' actions.

Chapter 9 revisits the research questions and discusses to what extent they have been answered.

[9] Figure 1.1 gives an alternative outline placing the chapters along a time line that also contains sketches of the prototypes used in the research process.

1.6 List of papers and collaborations

The chapters of thesis has been based upon a set of research papers and technical reports. Below the chapters with corresponding articles are grouped together.

Chapter 4, which describes the work with the Cero system, is based upon the following articles.

> *Anders Green and Kerstin Severinson Eklundh. Designing for Learnability in Human-Robot Communication. IEEE Transactions on Industrial Electronics, 50(4):644–650, 2003.*
>
> *Kerstin Severinson Eklundh, Anders Green, and Helge Hüttenrauch. Social and collaborative aspects of interaction with a service robot. Robotics and Autonomous Systems, 42(3–4):223–234, 2003. Special issue on Socially Interactive Robots.*
>
> *Helge Hüttenrauch, Anders Green, Mikael Norman, Lars Oestreicher, and Kerstin Severinson Eklundh. Involving Users in the Design of a Mobile Office Robot. Systems, Man and Cybernetics, Part C: Applications and reviews, 34(2):113–124, 2004.*

In Severinson Eklundh et al (2003) and Hüttenrauch et al (2004) my contribution was the sections that describe the spoken language user interface and the CERO character.

Chapters 4 and 5, which concern the elicitation of human-robot communication in a realistic use scenario using the Wizard-of-Oz method, is based on the following articles. In these articles my contributions concerned the discussion of the described methodological approach, the design of the user studies and data collection described in the paper was done in collaboration with Helge Hüttenrauch, Elin Anna Topp and Kerstin Severinson Eklundh.

> *Anders Green, Helge Hüttenrauch, and Kerstin Severinson Eklundh. Applying the Wizard-of-Oz framework to Cooperative Service Discovery and Configuration. In 13th IEEE International Workshop on Robot and Human Interactive Communication RO-MAN 2004, pages 575–580, 20-22 Sept 2004.*

Anders Green, Helge Hüttenrauch, and Elin Anna Topp. Measuring Up as an Intelligent Robot – On the Use of High-Fidelity Simulations for Human-Robot Interaction Research. In Proceedings of The 2006 Performance Metrics for Intelligent Systems Workshop, PerMIS'06, Gaithersburg, MD, USA, August 21-23 2006.

Chapter 5, which describes the collection and annotation of corpus material in the Cero and the Cogniron project is based on the following papers.

Anders Green, Helge Hüttenrauch, Elin Anna Topp, and Kerstin Severinson Eklundh. Developing a Contextualized Multimodal Corpus for Human-Robot Interaction. In Proceedings of the Fifth international conference on Language Resources and Evaluation LREC2006, 2006.

Nuno Otero, Anders Green, Chrystopher Nehaniv, Helge Hüttenrauch, Dag Syrdal, Kerstin Dautenhahn, and Kerstin Severinson Eklundh. Insights from corpora of embodied interaction with cognitive service robots. Technical report 472, School of Computer Science, University of Hertfordshire, 2007.

Anders Green, Helge Hüttenrauch and Kerstin Severinson Eklundh (2005). D1.3.1 report on the evaluation methodology of multi-modal dialogue. Technical report, COGNIRON. The Cognitive Robot Companion Integrated Project Information Society Technologies Priority, FP6-IST-002020.

My contribution was the development of annotation schemas for annotation of spoken and gestural data (in Otero et al. 2007) and the discussion of the corpus development in Green et al (2005; 2006a).

Chapter 6 treats miscommunication analysis and is based on the following article.

Anders Green, Britta Wrede, Kerstin Severinson Eklundh, and Shuyin Li. Integrating Miscommunication Analysis in the Natural Language Interface Design for a Service Robot. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems 2006 (IROS'06), pages 4678–4683, Beijing, China, October 9–15 2006.

Chapter 7 concerns how perception and design feedback could be incorporated in the design of human-robot communication and is based on the article:

> *Anders Green. The need for contact and perception feedback to support natural interactivity in human-robot communication. In Proceedings of IEEE 16th International Symposium on Robot and Human Interactive Communication (RO-MAN 2007), pages 552–557, Jeju, Korea, August 26-29 2007.*

My main contribution in this paper was the analysis of miscommunication, presented in the thesis, the design implications proposed in the paper were based on joint work with Britta Wrede and Shuyin Li.

Chapter 8 discusses spatiality and ways of designing with spatial prompts and is based on the following article. The term "spatial prompting" was coined by me, while the conceptualisation and analysis of it was joint work with Helge Hüttenrauch.

> *Anders Green and Helge Hüttenrauch. Making a Case for Spatial Prompting in Human-Robot Communication. In Multimodal Corpora: From Multimodal Behaviour theories to usable models, workshop at the Fifth international conference on Language Resources and Evaluation, LREC2006, Genova, Italy, May 22-27 2006.*

CHAPTER 2

Models and Design Approaches for Human-Robot Communication

In this chapter I will focus on how models of natural language use can be employed in the design of human-robot communication. I will do this by introducing some theoretical concepts of human-human communication that can be applied to human-robot communication in terms of *dialogue modeling*, *dialogue design* and *communicative quality*.

2.1 Human-Robot Communication as a situated activity

If we look at Human-Robot Communication in a broad perspective all instances of Human-Robot Interaction seem to involve communication to some degree. The focus in this thesis is on Human-Robot Communication as an activity where natural language is used to engage, manage and sustain joint activities. This involves the situated perception, understanding and expression of verbal, gestural and bodily signs.

To illustrate this in a use scenario we can consider Figure 2.1, which shows a situation from a scenario where a user teaches the name and location of an object to a robot. This is done in close proximity to the robot by means of verbal utterances and gestures used in combination. To handle this type of interaction in a computer based system it is necessary to perceive and interpret multimodal communicative actions that are displayed simultaneously. This include understanding verbal utterances, deixis through hand gestures as well as gaze. Apart from interpreting human communicative behaviour, the robot also needs to understand what is be-

U: Hello robot

R: I am ready

U: This is an orange

R: What is the object?

U: ...an orange

R: Found an orange

Figure 2.1 *In a human-robot scenario conversation unfolds both as verbal interchanges and gestures simultaneously.*

ing referenced. In the case of the example in Figure 2.1, the robot would need to understand what the referenced object is and how it should be distinguished from other objects. The robot also needs to be able to disambiguate the reference to that particular object with respect to other similar objects. Using natural language to disambiguate references to objects is one of the most interesting possible uses of natural language user interfaces in human-robot interaction.

Others have also noted that natural language communication is an important aspect of human-robot interaction. Klingspor et al (1997) characterise Human-Robot Communication as involving the following aspects, namely, providing instruction in an intuitive way, i.e, to "translate the user's intentions into correct and executable robot programs", and to provide feedback to the user so that she can understand what is happening on the robot's side (Klingspor et al., 1997). Communication with an embodied robot is also in focus in the definition of Human-Robot Interaction used by Hüttenrauch (2007): "the interaction and communication between a user and a mobile, physical robot". Communication is also proposed as an important factor for how socially interactive robots are perceived and accepted by humans (Fong et al., 2003b). Tenbrink (2003) points to the situatedness of human-robot communication and argues that robots need to be able to communicate about spatial features of the environment.

2.2 Cooperation, common ground and language use

Human to human natural language dialogue is affected by a set of factors ranging from physical and perceptual features of the participants, the semantic properties of the language in question, and perhaps most importantly, the social and cultural constraints on the situation in which the dialogue is carried out.

Cooperation

Cooperation on the basis of a shared understanding of the social conventions is an important feature of human language. Grice (1975) proposed that most conversations are carried out in a generally cooperative manner. This was captured in the CO-OPERATIVE PRINCIPLE, formulated as: "make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged". He also formulated sub-principles in the form of Maxims that further specifies the cooperative principle. In the section[1] on dialogue design I will discuss how Grice's Maxims can be applied to design of human-robot communication.

Common ground

Understanding one-another in order to collaborate, to co-ordinate joint tasks and to share experiences is essential for human communication (Allwood et al., 1991; Bunt, 1999; Clark, 1996; Goodwin, 2000). These approaches presented provide principles for conversation and are therefore useful both during analysis of interaction and when designing dialogue for natural language user interfaces. To achieve common ground during conversation humans engage in co-operative behaviour to achieve common goals (Allwood et al., 1991; Clark, 1996). The notion of *common ground* is used to describe a mutual process of sharing information between participants, *grounding*. In this process the interlocutors try to establish mutual belief by performing co-ordinated actions that are oriented towards a set of goals (Clark, 1996). It is also often assumed that the interlocutors are able to continuously monitor the actions and communicative behaviour of others (Clark and Krych, 2004).

In the grounding process the shared environment plays an important role by providing a *contextual configuration*, a set of locally relevant sign phenomena,

[1] See page 32

which Goodwin (2000) refers to as *semiotic field* instantiated in different media in the process of forming meaning and coordinating action as the interaction unfolds (Goodwin, 2000). The way communicative actions are understood depends on the preceding context as well as their ability to dynamically change the current context (Bunt, 1999).

Feedback

Another type of behaviour which is crucial for the communicative process in human-human conversation is feedback. Larsson (2003) defines feedback as "behaviour whose primary function is to deal with grounding of utterances in dialogue". Feedback can be categorised according to communicative function, which specifies what type of action is performed, and form, in which manner the feedback is displayed. The communicative function of feedback can be described in terms of the level of action, or basic communicative function, following Allwood (1995) and Clark (1996)[2]:

- Contact: feedback where the receiver, implicitly or explicitly, signal the willingness or ability to continue the communication. The interlocutors are in contact with each other.
- Perception: feedback concerning perception from the receiver signal whether the receiver, has perceived the utterance issued by the conversation partner.
- Understanding: is feedback that signals whether the receivers has understood the utterance from the interlocutor.
- Reaction: is feedback that addresses the main evocative intention of the interlocutor.

In addition to this, the feedback given in conversation can also have polarity, which may either negative, positive or neutral. Feedback may also be eliciting, meaning that it is aimed to evoke a response on the part the receiver (Allwood et al., 1991). In human-human conversation, verbal and body gestures can be seen as the two primary modes of producing feedback. Feedback signals may be either linguistic using back-channels or other linguistic structures like "m", "yes", "yeah", "sure"

[2]Clark (1996) used the terms: attention, identification, recognition and acceptance.

and tag questions, etc, or gestures like “nodding”, “shake ones head” or “raise eyebrows”, just to mention a few. Below are some examples of feedback:

(2)	**A:**	Hello!	
(3)	**B:**	OK I will talk to you	Explicit, Contact^{Positive}
(4)	**A:**	Hello!	
(5)	**B:**	⟨looks up⟩	Implicit, Contact^{Positive}
(6)	**A:**	What pages should I read?	
(7)	**B:**	Pages in what?	Explicit, Understanding^{Negative}
(8)	**A:**	I have sold my robot?	
(9)	**B:**	How much did you get?	Explicit, Reaction^{Neutral}

Grounding on the perceptual level

There is more to grounding than verbal and gestural feedback. Human perceptual behaviour plays an important role in establishing and creating meaning to arrive at common ground in conversation. The mutual experience of being *perceptually co-present* is triggered by salient event(s) that make people aware that they are sharing the same experience (Clark, 1996). Perceptually salient events may stem from communicative acts like speech, gestural indication and gaze, as well as from partner activities and other perceivable events (like a telephone ringing).

From a psychological point of view, salience of a perceptual event, such as the occurrence of human speech, is determined by its relative strength dependent on context and stimuli. Events with a high relative strength are most likely to draw our attention (Pashler et al., 2001). But our goal-oriented behaviour also affects the ability to perceive stimuli. This means that we are more or less attuned to stimuli of a particular kind, depending on our current activities. This suggests that cognitive processes allow humans to actively focus perceptual attention (Cherry, 1953).

Gaze

One type of perceptual events that are especially important in human-robot interaction is human gaze. Psychologists generally agree that humans have modular perceptual subsystems for recognising gaze directions (Langton et al., 2000; Wilson et al., 2000). It is well known that gaze has a strong salience and that people have a good discrimination of the line of gaze of others (Gibson and Pick, 1963).

The direction in which another person is looking gives important cues to the focus of attention of the person, something which is important in collaborative settings, for instance, when monitoring the actions of other participants (Clark and Krych, 2004). Interpretation of gesture and human activity, including the gaze of others has been studied in different ways in Human-Robot Interaction contexts, for instance by Sidner et al (2005) who studied the role of gaze to establish the degree of engagement of users in human-robot communication. Torrey et al (2007) investigated whether a robotic system could increase its responsiveness by adapting to the user while monitoring the user's gaze and delays in the task progress.

Gesture

Multimodal interaction provides a challenge for dialogue research, since it involves information that is not easily described using a formal model. This is especially obvious in the case of gesture research. Human gesture and body language have been studied by a number of researchers with various goals. Theories of gestures have been used to account for pragmatic meaning, primary applied to conversation (like Clark and Krych 2004; Gill et al. 1999, 2000; Kendon 1997), and to investigate psychological phenomena (cf. Ekman and Friesen 1969; McNeill 1992).

Ekman and Friesen (1969) categorise gestures departing from work that was discussed by Efron (1972). Their categorisation was mainly descriptive. Another type of categorisation, mainly focused on narrative gesture, was used by McNeill (1992). Mcneill's taxonomy has been used in computational approaches for recognising narrative structure in discourse (Quek et al., 2000).

Kendon (1997) proposed the following categorisation to account for the ways in which meaning can be formed by using gesture and speech. Gestures can either be used *alone* or *co-produced* with speech. The components of a gesture may contribute to the meaning of an utterance in several aspects:

The gesture may provide *content*, by which meaning is emphasised or influenced depending on the meaning of the utterance. Another aspect is *deixis* by which a reference to a domain object is made. Gestures may also be produced alongside speech, as *conjunct* gestures, that do not provide lexical meaning (for instance, gesticulation alongside intonational patterns).

Communicative functions of the body

When it comes to understanding gesture to account for interactive communication there have been few attempts that incorporate an analysis of gestures viewing them as having conversational functions. Gill (1999) extends the framework of dialogue moves (Carletta et al., 1997) to include the notion of body moves. Gill (1999) does not classify the kinetic movements of the body, instead she focuses on the functional aspect of the gesture. A body move may be a response to another body move or a verbal utterance. The notion of body moves is broader than specific conventional speech acts or dialogue moves.

A body move might be multifunctional, for instance, whereas a verbal utterance like *"yes"* usually has a single[3] function (like acceptance), a body move may also at the same time create a sense of contact. Gill et al (2000) refer to this as a *space of engagement* between the participants in a conversation. The notion of an engagement space has been discussed in several theories that focus on the management of space. Kendon (1990) studied spatial configurations and describes the relation when two participants have a common perceptual focus as an *o-space*, or *transactional space* which is located in an area which is perceptually mutually available of the participants. A typical[4] configuration is in the visually shared environment between two participants that are facing eachother. It is within this area that interaction is conducted. Clark (2004) refers to interaction space as the *workspace*, where perceptual co-presence is established between speakers (Clark, 1996; Clark and Krych, 2004).

Gill's notion of Body moves is interesting since it relates gestures to communicative theories that include the understanding of perceptual and attentional status of the participants. In these theories communication is viewed as a shared activity between participants. I have already mentioned Goodwin's (2000) account of situated communication where interaction is seen as an activity that involves the use of the whole body and the surrounding context as a backdrop for the unfolding interaction. Clark and Krych (2004) stress the importance of providing a bilateral account to model human-human communication. In such an theory it is important to describe and explain how the communicative status of one another is communi-

[3] Verbal utterances may be multifunctional, too.

[4] Other ways of negotiating transactional space is indeed possible, for instance, by using touch and audio.

cated. Here the display of feedback regarding perceptual status and willingness to interact plays an important role (Allwood, 2002; Bunt, 2000).

Gill (1999) categorises body movements[5] that are used to display communicative status along the following dimensions:

- Referencing: which is used to indicate or demonstrate a reference to a situation, like directing the body towards an object.
- Contact and communicative attitude: which is used to intitiate or display attitudes towards the willingness to continue interaction, for instance by turning towards the conversation partner.
- Focusing: the act of transferring attention to a certain physical or abstract spot in the situation, for instance, by placing the body on a specific point in the engagement space to indicate a new point of interest.

Focusing is especially interesting since it concerns the engagement space (o-space or workspace). Focusing using the whole body, is a kind of deixis, but according to Gill (2000) it also provides a meta-discursive function that signals a shift in the center of attention in the discussion, like a shift in body posture with the same meaning as the utterance "I am going to focus on this spot". Projected change is important in Schegloff's (1998) notion of *body torque*. Body torque is a state of the bodily configuration when two different body segments are oriented in different directions. The unstable configuration of the body "projects change", meaning that the participants may predict that a shift in posture is pending. For instance, when turning the head towards something, this might predict a change of the general body orientation and consequently a new configuration of engagement space.

2.3 Natural language dialogue modeling

Human-to-human dialogue can be viewed from different perspectives. The phenomena modeled by researchers studying dialogue occur on different levels. On the sentence level, models that employ Speech Acts (Austin, 1962; Searle, 1969) are used to account for the semantic and pragmatic meaning of utterances.

[5] See Allwood (2002) for an extensive account of means of producing communicative functions using the human body.

Adjacency pairs

The way speech acts capture the propositional content fits very well with approaches that represent dialogues in shallow structures, such as adjacency pairs, that are formed by an initiative and a response (Levinson, 1983). The constituents of an initiative-response structure pair can be analysed with respect to their communicative function, for instance, QUESTION–ANSWER and can be used to analyse interchanges between humans and robots, such as this one, taken from the corpus described in Chapter 5:

(10) **R:** Is this the object? (Question)

(11) **U:** Yes, it is the object. (Answer)

Adjacency pairs identified in other dialogue domains can in principle be used to capture general dialogue phenomena. In a robotics scenario, an adjacency pair SUMMONS–ANSWER which is typical for telephone conversations (Schegloff, 1979) can also be used to capture initialisations of conversations with robots:

(12) **U:** Robot! (Summons)

(13) **R:** Hello, I am ready. (Answer)

The notion of adjacency pairs has been influential for practical approaches for building dialogue systems (cf. Ahrenberg et al. 1990). By analysing dialogue using structural relations based on adjacency pairs, interaction situations that are limited to a single modality can be handled, such as telephony based systems for time-table information.

One phenomenon which can be modelled using adjacency pairs or in terms of local communicative functions is conversational feedback. Feedback provides one of the most important resources for enabling the grounding process in dialogue. Speech and body gestures can be viewed as the primary modes of production of feedback. Feedback signals may either be linguistic, using back-channels and other linguistic structures, or non-verbal using the body to issue gestures (gestures like nod, shake head, raise eyebrows). This means that speech and body gestures can be viewed as the two primary modes of production of feedback. These two modalities either reinforce each other by introducing redundancy or add information to one another (Allwood, 2002).

The model proposed by Traum (1996) accounts for conversational acts and the way they change the beliefs of participants in dialogue. To do this it is suggested

that we need to handle units that are smaller than the sentence level to capture dialogue. Traum et al (1994; 1996) proposed a model that describes dialogue functions for partial sentences, *Utterance-Units*[6] By analysing functions of utterance units, rather than whole sentences, dialogue phenomena can be handled on two different levels: Feedback and turn-taking acts used to manage the dialogue are associated to sub-utterance units (e.g., repairs, acknowledgements and initiations). Grounding acts, concerning the topic of conversation, are associated with core speech acts (e.g. inform, questions, answer, etc). In the example below the utterance "is this the object" concerns the core task, i.e., negotiating the character of objects in the environment. The utterance "yes" is treated as an utterance unit with the function of providing positive feedback:

(14)	**R**	is this the object	(Question)
(15)	**U**	yes $\langle \ldots \rangle$	(Feedback+)
(16)	**U**	...it is the object	(Answer)

Plan-based approaches

Binary relations, such as adjacency pairs, fail to represent the more complex dialogue phenomena that are needed to model a more natural style of conversation. In plan-based approaches, dialogue models are used to represent the underlying planning that gives rise to dialogue contributions. Grosz and Sidner (1986) approached dialogue from the perspective that the *intentions* of the participants also need to be represented in order to handle dialogue. They represented this as a trifold relationship between components that are dependent upon each other: the *linguistic structure* of the sequence of utterances in the dialogue, the *structure of intentions* and an *attentional state*. On the utterance level *communicative functions* are aggregated into *discourse segments* that account for the sequencing of utterances. On an abstract level, that concerns the overall purpose of the conversation, *discourse purposes*, are used to model the intentions of the participants, which can be seen as interpretations for why the specific discourse acts are being performed

The discourse segments and the discourse purposes are structured with respect to the participants (verbal) focus of attention using the notion of an *attentional*

[6]Utterance-units are defined as continuous speech by the same speaker, punctuated by prosodic boundaries (i.e. pauses, boundary tones) making it possible to split the utterances into utterances units by algorithmic means.

stack which provides a representation of the discourse. The representation is a dominance hierarchy of discourse purposes that determines the structure of the dialogue (Grosz and Sidner, 1986).

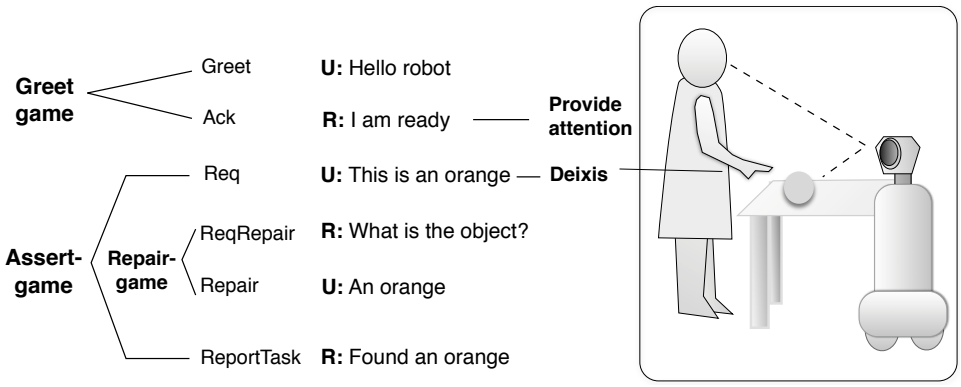


Figure 2.2 *The scene from Figure 2.1 expressed as a dialogue game, together with multimodal conversational acts Provide-Attention and Deixis (or Reference).*

Dialogue games

Another way of describing dialogue is based upon the notion of conversational games (Power, 1979), or a dialogue game. A dialogue game can be described by using a dialogue model based on utterance function and game structure (Carletta et al., 1997; Kowtko et al., 1991). In a dialogue game participants engage in conversation where rules are determined depending on the character of the game. The notion of dialogue games has been used within artificial intelligence. One example is Power (1979) who let *virtual robots* engage in dialogue games in a *blocks world*. When robots in the blocks world successfully performed a dialogue game it led to a change of state in the world or in the participating robots. For instance, a successful FIND_OUT game, lead to increased information, whereas a GET_DONE game incited a (partner) robot to perform an action in the world (Kowtko et al., 1991; Power, 1979).

Dialogue games can be used to conceptualise human-robot interaction in real-world scenarios. The notion of dialogue games stem from the theory of Sinclair and Coulthard (1975). Their model was strict in the sense that they used a hi-

erarchy where at the highest level, dialogue games form *transactions* made up of *exchanges*, that are made up by *moves*, and at the lower level, *acts*, roughly corresponding to the speech acts of (Searle, 1969). As noted by Severinson Eklundh (1983) the levels of Sinclair and Coulthard (1975) do not suffice as sometimes meta notation is used in their examples to mark that categories of two levels are related, meaning that they need an extra level of description. More recent theories, like Carletta's (1997) allow for dialogue games to be structurally embedded. In Figure 2.2 a dialogue game representation of a sequence of human-robot interaction is depicted. The model represents dialogue structure on three levels. At the lowest level dialogue is modelled using moves corresponding to a speech act. At the next level, a dialogue game, is formed out of a set of utterances starting with an initiation, encompassing all utterances up until a certain purpose of the game has been either fulfilled or abandoned. Games are themselves made up of conversational moves, which are simply different kinds of initiations and responses classified according to their purposes (Carletta et al., 1997). As we mentioned above, dialogue games may have other games as embedded structures (Carletta et al., 1997; Severinson Eklundh, 1983). This is depicted in Figure 2.2, where a Repair game is embedded in a Goto-game. At the lowest level *moves* roughly corresponds to the notion of speech acts.

Incorporation of Mental models

Other approaches for modeling dialogue which are inspired in the notion of grounding, models the mental state of the human conversation partner. The BDI-model (*Beliefs, Desires, Intentions*) was used by Traum and Allen (1995; 1994). They provided a computational model of grounding that comprised formal rules for grounding. Their application was task-oriented acts of argumentation for route planning (of trains), in a virtual world. The actions of agents in the corpora and systems studied are performed either in an abstract information seeking task or in a virtual environment. Information management is also considered by Larsson (2002) who models dialogue in terms of issues, i.e., information that is useful for some activity. The issues are semantically modelled as questions which the system has to address rather than identifying plans as in the approach by used by Traum and Allen (1995; 1994). This approach has been used for modeling human-robot dialogue in the Carl system (Quinderé et al., 2007).

Models that incorporate Context

The ability to account for context change is very important in human-robot interaction, since participants in conversation both can change the context actively by performing speech acts as well as through physical acts. In human-human scenarios Goodwin (2000) has described this qualitatively as an ongoing semiotic process. Bunt (2000) provides a more detailed and formal account of context change. In Dynamic Interpretation Theory he stressed the context as an important factor for distinguishing different functions within dialogue. Bunt considers *perception* and *interpretation* acts. In his view *dialogue acts* are *"the functional units used by the speaker to change the context"*. Similarly to the model used by Traum (1996) Bunt (2000) divides dialogue acts into *task-oriented* acts and acts that are used for *dialogue-control* (e.g., feedback). The distinguishing feature of these is that while task-oriented acts change the semantic context, the dialog-control acts change the social and physical context but do not affect the semantic context (Bunt, 2000).

2.4 Dialogue design guidelines

Dialogue design can be approached from two perspectives. Partly it is a creative activity which can be approached from a practical perspective, and partly it can be viewed as a way of linking between abstract dialogue models and robot actions. Very few attempts have been made to establish guidelines for *multi-modal interfaces* for service robots. This stems from the fact that human-robot interaction is a relatively new discipline lacking cases of actual product development intended for end users. For spoken dialogue systems there are several approaches, either they are *principled*, deriving from theories for communicative quality or they are more *practically oriented* based on experiences of dialogue design.

Practically oriented guidelines

A good starting point for a discussion on practically oriented guidelines is provided by Dybkjær et al (1998; 1997) who discuss some challenges related to practical design of spoken dialogue systems. Their focus on applications means that there is a need to focus on users that bring real tasks to the system. This means that developers need a solid understanding of what type of service to what type of users

that is going to be provided. There are also technical and pragmatic considerations that have to be considered when it comes to the design of the practical system:

- The quality of speech recognisers and the linguistics analysis.
- The output voice quality, whether to use prerecorded speech or synthesised speech.
- Ways of providing relevant feedback using appropriate phrasing.
- Appropriate dialogue models and initiative management.
- Adequate error handling, something which is coupled with providing help or interaction guidance.

The first two challenges related to speech input and output are partly a matter of acquiring state-of-the art software and hardware and partly a scientific problem which falls out of scope of this thesis. If we turn to ways of providing relevant feedback and to manage dialogue, the practically oriented approaches concern *prompt design*, i.e., ways of designing the output of the system so that it influences the conversational behaviour of its users. A set of practical techniques to guide the users' speech are discussed by Yankelovich (1996). Prompt design is important in order arrive at a usable system. Phrasing is essential but dialogue management strategies and natural language understanding are also important when it comes to decide what to say. To illustrate how utterances may be phrased on the surface level I have used the categorisation by Yankelovich (1996) to construct a set of examples that are oriented towards the service robot domain, shown in Table 2.1.

Another practically oriented approach is offered by the framework called *Universal Speech Interfaces* (USI) described by Tomko et al (2005). The fundamental hypothesis of USI is that when a user has acquired the skills needed to handled one USI-based application, learning speed is improved for new applications that use the same interface model. This is approached by constructing a query language called *Speech Graffiti* inspired by the way graphical user interfaces and text input are becoming conventionalised and can serve as a standard way of providing an interface for heterogeneous types of applications. Query languages for that are intended to work for different applications are using the same commands and style of interaction (i.e., the "say-and-sound" of the interface) for the types of phenomena recurring in many dialogues e.g., asking for help ("where am I"), error handling

Table 2.1 *Prompt types.*

Explicit prompts	*Deliver an object. Please **say the name** of the object and to where it should be delivered!*
Implicit prompts	*Go, **where** do you want to go?*
Incremental prompts	*Deliver, what do you want to deliver...* *⟨silence⟩* *...say an object.*
Tapering	*(First interaction) Deliver. **Say an object** and a location!*
	*(Second interaction) Deliver. **Specify an object to be delivered** and a location where it should be delivered!*
Hints	*You can say follow, and then move away slowly to make me start following!*

("scratch that", "start over"), navigation commands ("more", "next", "previous", etc). To handle the domain specific tasks phrases like "what is ⟨domain keyword⟩" (data base query) and "go ahead" (send to application) are used together with key-phrases. The USI framework was originally developed for the domain to information access, for instance to search in a movie database. Recent developments has extended the framework to household appliances, such as video-camcorders, home audio and video equipment and software media players (Nichols et al., 2003).

Principled guidelines

In recent years different attempts at providing guidelines for designing spoken language interfaces based on communicative principles have been introduced. One prominent example are the guidelines by Bernsen and Dybkjær (1998). They have used the maxims of Grice (1975) to motivate guidelines for spoken language user interfaces consisting of about 25 generic and specific principles that can be used to design and evaluate usability of dialogue systems. In the following will present some of the specific principles that are relevant for this thesis[7]. They proposed seven general aspects of interaction and for each aspect some specific design principles. Some of these principles have the same wording as the maxims of Grice,

[7]The book of Bernsen and Dybkjær (1998) provides an in-depth overview and motivation.

and some are specific to dialogue systems development. The first four aspects proposed by Bernsen and Dybkjer (1998) directly correspond to, or use the same wording as Grice's Maxims:

- *Informativeness: "Make your contribution as informative as is required (for the current purposes of the exchange)."*(Maxim of Quantity).
- *Truth and evidence: "Do not say that for which you lack adequate evidence"*, (Maxim of Quality).
- *Relevance: "Be relevant, i.e. Be appropriate to the immediate needs at each stage of the transaction."*, (Maxim of Relation).
- *Manner: "Avoid obscurity of expression", "Avoid ambiguity", "Be brief", "Be orderly"*, (Maxim of Manner).

In addition to these dialogue aspects, that can be explained in terms of Grice's Maxims, Bernsen and Dybkjer (1998) also formulates principles that are specific to spoken dialogue user interfaces:

- *Partner asymmetry:* "Inform the dialogue partners of important non-normal characteristics which they should take into account in order to behave cooperatively in dialogue". This principle has to do with situations where one or more dialogue partners are not in a normal condition or situation.
- *Background knowledge:* "Take users' background knowledge into account". This principle has to do with what users know before starting the dialogue or what the user learns during the dialogue.
- *Repair and clarification:* "Provide ability to initiate repair if system understanding has failed".

The generic principles also subsume specific dialogue principles, formulated by Bernsen and Dybkjer (1998), who are targeted to dialogue design. The generic (gricean) principle related to informativeness subsumes the two specific principles: *"Be fully explicit in communicating to users the commitments they have made"* and *"Provide feedback on each piece of information provided by the user"*. With respect to Manner the specific principle *"Provide same formulation*[8] *of the same*

[8][phrase or expression.]

question (or address) to users everywhere in the system dialogue turns", can be seen as reflecting the goal of providing consistency in user interfaces.

Gamm and Haeb-Umback (1995) propose a set of guidelines for interfaces to consumer electronics. Some of these are directly subsumed by the guidelines proposed by Bernsen and Dybkjær (1998), for instance those concerning consistency and feedback. Two of them are worth examining a bit closer, namely *"Give the user the choice of input modality"* and *"Do not overload the voice input channel"*. These guidelines are related to what Dybkjær and Bernsen (2000) would consider *modality appropriateness*, i.e., to what extent a specific interface modality is appropriate with respect to a particular domain task.

Rosset et al (1999) discuss design aspects that should to be taken into account when designing spoken language user interfaces, termed *Ergonomic Choices*. The design aspects are not formulated as guidelines in the classical sense, instead they can be thought of as topics to consider or goals to strive for in the design process. Following Rosset et al (1999) these design aspects can be phrased as:

- *Freedom and flexibility:* Avoid imposing constraints as long as the dialogue flows well,
- *Negotiation:* Provide the possibility to accept or refuse system proposals, and,
- *Navigation:* Support identification of a change in the task.

These decisions are in line with similar approaches and guidelines. The guidelines proposed by Bernsen et al (Bernsen et al., 1998) are more detailed. Rosset (1999) also discuss two areas that go beyond concerns that have to do with information content. These areas for which decisions have to be made are concerned with the flow and timing of interaction and the management of initiative during interaction. Following Rosset et al (1999) these design aspects can be phrased as:

- *Initiative:* Handle how the progression of the dialogue is directed by providing a mixed-initiative dialogue handling strategy, and,
- *Contact:* Never let the user get lost and provide immediate response when addressed by the user.

Handling vocabulary challenges

Several challenges of design can be seen in relation to communicative quality and dialogue design are related to *habitability*, something which may be defined as a relation between what the system can manage and what the user feels comfortable with. A habitable interface is one where the users do not feel unnecessarily constrained by the systems ability of understanding (Hone and Baber, 2001). We learn and adapt to patterns of interaction that are common to all humans within a certain community. For this purpose we possess mental models, just like we have mental models of interaction with doorknobs, stoves and light switches.

Providing a full-blown natural interaction giving the user maximum freedom of expression (Rosset et al., 1999) means that the designer faces a combinatorial explosion of possible phrases. There is a trade-off between the number of possible user utterances and the lexicon and grammars employed in a system. In language the mapping of terms to referents is many-to-one, this problem is termed the vocabulary problem and needs to be addressed in any practical system (Brennan, 2001). There have been some attempts at fighting the combinatorial explosion. Identifying a restricted subset of the particular language by looking at frequencies may be a way to overcome some of the problems. Even if this approach looks promising from a theoretical point of view, there are empirical counter-claims (Brennan, 2001). One way of overcoming the vocabulary problem is to use careful phrasing in system responses. The psychological phenomenon which is used is referred to as lexical entrainment and means that the user's language is coloured by the way the computer speaks, meaning that users adapt to the phrasing used by the computer (Bell, 2003; Brennan, 2001; Zoltan-Ford, 1991). This mechanism may be used to control the behaviour of the user so that linguistic variability decreases. Lexical entrainment has been investigated and actively used in design in practical systems, for instance, by designing closed questions to limit the range of input (Rosenfeld et al., 2000; Yankelovich, 1996). Another way of limiting the search space is to be aware that lexical entrainment can be used to reduce the size of the lexicon that needs to be active during dialogues with specific user, using the circumstance that variability is high between different dialogues with different users but relatively low within a single dialogue (Brennan, 2001).

2.5 Design for Human-Robot Communication

Designing human-robot communication involves the creation of a system that responds to communicative actions of its users. The way the system responds to user actions decides to what extent the system is successfully engaging in communication with its users. To provide services a robot performs physical and virtual actions, for instance learning things about the environment. A service robot can perform physical and virtual actions without engaging in communication with its users, i.e., by acting autonomously. In an interactive system communicative actions provides the way to interface the robot's underlying task capability and are used to establish, maintain, and influence the physical and virtual actions that form the services provided by the robot. Communicative actions can also be included in the range of possible services[9] that a robot system provides. This relationship is depicted in Figure 2.3.

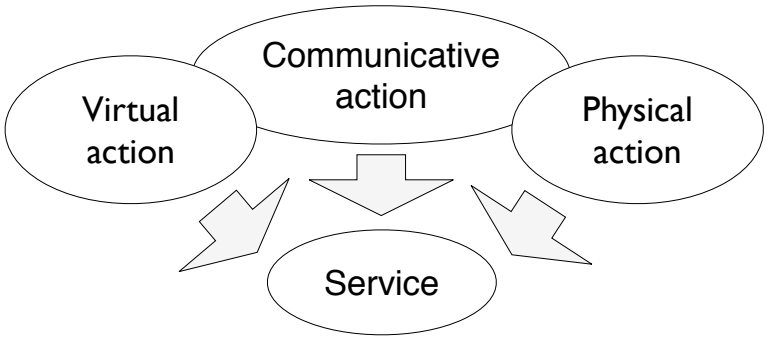


Figure 2.3 *Providing services through, communicative, physical and virtual actions*

When designing communication for a robot we work within a design space that includes the whole range of language technology that corresponds to, or mimic, human communicative capability, including speech recognition and synthesis, dialogue management, gesture interpretation and generation, face-detection, detection and recognition of human activities, etc. We also need to consider less complex sounds, movements, shapes and last but not least all attributes that are possible to

[9] An example of a robot system that is primarily oriented towards interaction is RoboVie, described by Kanda et al (2002b).

use by the introduction of classical user interface components used for computers, like windows, icons, menus, pointing devices etc. To these interaction modalities we should add physical system actions, such as spatial positioning and the physical service tasks performed by the system.

When examining the research approaches for human-robot communication two patterns emerge:

- PHYSICAL TASK FOCUS, ROBOTS AS APPLIANCES, meaning machines that are designed to perform specific services in the home or the workplace. These services are related to handling objects or dependent on the robot being situated in a specific environment.
- INTERACTION FOCUS, ROBOTS AS CREATURES OR CHARACTERS, referring to robots designed as humanoids or animals with the intent of being socially interactive. Practical tasks for these robots are possible, but related to management of information rather than manipulation of the environment.

These perspectives are perhaps better seen as two extremes on a varying scale. A robot, like the fetch-and-carry robot MOPS Tschichold-Gürman et al. 1999, which was designed with the goal of solving physical service tasks rather than being an artificial character is an example of a robot with a task-focus. An anthropomorphic robot, like Robovie (Kanda et al., 2002a), equipped with eyes, mouth, arms and legs and whose primary goal is to engage in conversation, is an example of a system created with an interaction focus in mind.

Integrated robot interfaces

One type of robot which clearly does not have externally mounted interface components are humanoids, such as the human-like androids at ATR described by (Ishiguro and Minato, 2005). The way humanoid robots have appeared as technical demos at exhibitions and science fairs make it hard for us to analyse their interactive capabilities in terms of whether they perform physical tasks or are intended for interaction. To the extent that humanoids are interactive it is instead another characteristic which is interesting, namely the integration of interface and embodiment.

The interaction modalities that have been put forth as the primary means for supporting interactive behaviours for robots can be grouped into two main cate-

gories. One approach involve robots that have modalities that use a conventional human-computer interaction style of interaction. In this group we find graphical user interfaces, handheld computers, touchscreens and button.

The other approach is to use modalities that emulate or display human-like characteristics or capabilities in some way or the other, for instance by analysing or generating natural language. In an emulative approach robot design is supposed to match humans behaviour, for instance by equipping the robot with human-like equivalents to a body and a face used for gesturing, vocal organs for speech, etc. The most obvious example of this are biomorphic robots, referring to robots that have been given an anthropomorphic and/or animal-like shape.

Design of robots with natural language capability has been approached in two primary ways. One way involves robots that are interfaced using an intermediate device of some sort that provides an interface model which is independent of the robot. Interaction is then carried out through the device which act as a link between the robot and the user. The other way means that the robot acts in its own capacity, meaning that *"the robot is the interface"* (Hüttenrauch, 2007). This dichotomy has some consequences on the way we view and study human-robot communication. Designing interfaces for robots that are physically or conceptually detached from the robot provides interface challenges for which approaches of human-computer interaction design are applicable. There are numerous examples of interfaces where the robot is controlled using an interface hosted on separate device, speech dialogue and a graphic based on a desktop computer (Lemon et al., 2002), speech dialogue and multimodal input on handheld device (Perzanowski et al., 2003, 2001) or a handheld device with interactive menus (Fong et al., 2001, 2003b; Hüttenrauch and Norman, 2001). Whether the external control interface is using natural language as the bearing interface metaphor is not an issue here, but has consequences for the interaction situation as the robot and the user are not necessarily situated together.

Some communicative functions for service robots

Descriptions and categorisations of communicative functions that are accommodated in dialogue systems for service robots are not very common in literature. The types of goals and communicative actions that can be given as input to a robot system are heavily dependent on the scenario in which the system is going to op-

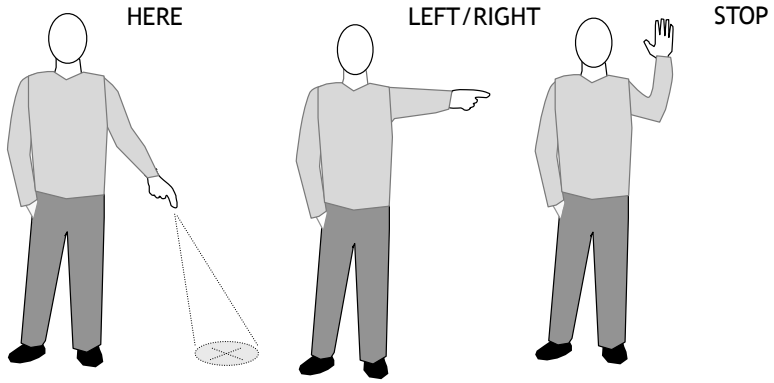


Figure 2.4 *Examples of three possible gestures that can be provided as input to a service robot.*

erate. The list below is an attempt to summarise and give a characterisation of the relevant communicative functions:

- Action-directives
  - Directive (short-term) goals. Utterances of this type define system actions to be performed immediately (within milliseconds, seconds). For instance, verbal utterances like "go forward", "stop", "pick red ball" or gestures. Gestures might be emblems like "⟨STOP-GESTURE⟩" with the same meaning as the spoken utterance "stop" or "⟨LEFT⟩" to specify which direction[10] to go (see Figure 2.4).
  - Protractive (long-term) goal. Utterances of this type define system actions with goals that are held for a long time (within minutes, hours, days) and possibly stretching over long distances[11], for instance the corridors of an office: "Go to Mary's office", "Guard this area" or "Find John".
- Information-requests, usually in the form of questions concern the information state of the robot rather than its physical task capability. For instance, "what is the time?", "Have you delivered the coffee?", etc.

[10] This gesture can (partially) be understood as a deictic gesture, or as having a deictic component

[11] Torrance (1994) uses the term long-*range* goals. I prefer to stress the temporal aspect of goals rather than the distal.

- Assertions concerning some state of affairs with the intention of updating the robot's knowledge in some respect. For instance "This is a my favourite book", "there is a large room at the end of the corridor", "hello robot, my name is Anders" or "I do not like coffee".
- Meta-communicative actions[12] are meant to provide information on or eliciting actions related to communicative aspects of interaction, such as grounding-acts (Bunt, 2000; Clark and Schaefer, 1989; Traum, 1994) or acts that are used to change the communicative context: "What did you say?", "I can't see!", gestures such as feedback-requests. Gestures may be emblems like ⟨SHRUGGING⟩, to signal that the other party does not understand, body moves proposed by (Gill et al., 2000) that have a meta-conversational function, or gestures that accompany speech like *beats* (McNeill, 1992).

For robots designs that that try to mimic human behaviour, this list should in principle coincide with any general theory on human-human natural language interaction that provides an account of communicative actions.

Task type and grounding strategies

McTear (2002) categorises natural language dialogue systems into state-based, frame-based, and agent-based based on their general control strategy. This overall control strategy determines what type of input is possible to give to the system, what means of verification the system has, the dialogue model that is used to implement the control strategy, and to what extent the system comprises a user model. Typically a state-based system comprises a very shallow user model and provides verification to the user through explicit confirmation of input. In a frame-based system the dialogue model can be represented using an information state, and the verification of input is explicit or implicit based on the frame that is activated. In an agent-based system the goal is to accommodate unrestricted natural language capability. These systems typically model the user's belief and intentions in order to provide verification using grounding as a model.

State- and frame-based models are commonly used for dialogue handling in natural language user interfaces to service robots. Systems that implement a dia-

[12] This category can be refined further.

logue agent as a control strategy are less common even if the robot itself appears as an biomorphic or anthropomorphic character.

Command-based interfaces are usually implemented using a state-based control strategy. Systems where the user writes or speaks single phrases and the robot responds by performing actions, like moving or picking up things, perhaps represent the simplest and most direct form of human-robot interaction. Zelek (1997) investigated the possibilities for a controlled language to control an autonomous robot. His structure was based on action verbs, GO and FIND which could be parametrised with destination, direction and speed. The grammar below, quoted from Zelek (1997), describes the structured language that can be used to give navigational commands to the robot:

$$\text{COMMAND} = \begin{array}{c} \text{VERB} \\ (\textit{from})\ \text{SOURCE} \\ (\textit{to})\ \text{DESTINATION} \\ \text{DIRECTION} \\ \text{SPEED} \end{array}$$

In Zelek's system the input to the system consists solely of action-directives (Zelek, 1997). The robot's communicative responses consisted of physical actions rather than linguistic actions.

Tellex and Roy (2006) describe a system that handles commands for activating spatial actions of the robot. The system of Tellex and Roy (2006) also handles commands similar to that of Zelek (1997) by grounding lexical functions in spatial routines for the system, but extends Zelek's approach by taking the situation into consideration e.g., "go right" means different things when being positioned at an intersection (i.e., "go to the intersection and turn right") versus being positioned on an open area (i.e, "turn 90 degree right and then go forward").

Both Zelek's system and the system of Telex and Roy (2006) provides responses in the form of physical actions, without verbal feedback. The physical actions of the robot can be viewed as an *evidence of understanding* (Clark, 1996; Skantze, 2007).

Even if Tellex and Roy (2006) used a state-base dialogue model, more complex models are needed as the domain task becomes more complex. This is evident in the system described by Skubic et al (2004) who use spatial information derived

from sensor input to generate linguistic descriptions. This provides the means for the robot to engage in a kind of meta-dialogue regarding the spatial context rather than controlling the systems movement. The system is able to engage in dialogue that is closer to conversation than with the approaches discussed previously, where physical action resulted from a successful issuing of a command. The system handles questions regarding the spatial context of the robot, for instance: *"Where is the nearest object on your left?"*. Questions of this type is then answered by providing a qualitative description of the spatial context of the robot: *"The object #1 is mostly in front of me but somewhat to the left. The object is close."* (Skubic et al., 2004). But even if this system provides verbalised communicative responses, the overall dialogue control strategy is still state-based.

The dialogue system used in the JIJO-2 system uses a frame-based dialogue model (Asoh et al., 2001; Matsui et al., 1997, 1999). JIJO-2 is an office robot performing dialogue in Japanese. The system recognises interrogative and declarative statements (e.g. *"go to matsui"*, *"where is matsui"*) to access database information. The input is interpreted using a task-frame and system knowledge. The robot behaviour and the output that control the flow of dialogue is provided using a template based on the task-frames.

An important difference between the Jijo-2 system and the systems described by Zelek (1997), Tellex (2006) and Skubic et al (2004) is that the robot is intended to *carry out tasks* in a work environment whereas the state-based systems are used to explore linguistic phenomena (command structure, spatial expressions, etc) in a one room lab area (Skubic et al., 2004; Zelek, 1997) or a simulation (Tellex and Roy, 2006). The need for a more cautious grounding strategy (Larsson, 2002, p. 97) becomes clear in a system that performs physical actions or actions that are irreversible.

In the Jijo-2 system (2001; 1999) imperative statements are used for issuing motion and task commands. From a grounding point of view in the system the way acts are grounded is carried out similarly both for information oriented tasks (e.g., updating a database) or movements, carried out in the physical environment (e.g., to show the way to an office). The use of a cautious strategy in a dialogue system means that after a command has been issued the dialogue system asks for confirmation before executing the task. This is shown in the following example

where the robot requests confirmation before executing the action of sending an e-mail:

(17) **U1:** Asoh-san wa doko?
(Where is Mr. Asoh?)

(18) **R:** Honkan ni imasu.
([He is] in the main building.)

(19) **U1:** Yonde kudasai.
(Please call [him].)

(20) **R:** Asoh-san ni me-ru wo okurimashouka?
(Shall I send an email to Mr. Asoh?)

(21) **U1:** Hai.
(Yes, please.)
⟨ *The robot sends an email* ⟩

Example 2.5.1: Frame-based, information-oriented dialogue with JIJO-2. Example from Asoh et al (2001)

For a physical task, the cautious strategy used is different, like in the following example:

(23) **U1:** Hara-san wa?
([Where is] Dr. Hara?)

(24) **R:** Office ni imasu.
([He is] in [his] office.)

(25) **U1:** Tsureteitte kudasai.
(Please take me [to his office].)
⟨ *starts moving* ⟩

Example 2.5.2: A dialogue that results in a physical task (Asoh et al., 2001)

In the example above the phrase "Tsureteitte kudasai" (*Please take me [to his office]*) is an explicit request to carry out a previously specified action. Once this phrase has been uttered the system starts to move.

In this section I have described some of the considerations that have to be made when designing dialogue for a system that affect the physical and spatial context. This can be summarised as follows:

- Robot actions can be used directly as an evidence of understanding depending on whether or not the use context allows it, e.g., if the consequences of physical or information actions can be considered safe.
- A cautious grounding strategy is used for tasks that have effects on the physical environment.

2.6 Chapter summary

This chapter has introduced Human-Robot Communication as situated activity, where natural language is used as the primary means of conveying information. Cooperation between participants is essential to achieve a successful communication in human-human situations.

- To handle dialogue in a computer it is necessary to model complex structures, for instance by using the notion of language games together with a model of the user's mental state. It is also important to take communicative effects that change the context into consideration.
- Using the notion of grounding provides some of the theoretical means to analyse and model the communicative process.
- Grounding concerns the establishing of mutual information related to the situated context of the user and robot through communicative action, including verbal and bodily communication of both parties.
- Phenomena related to perception is important for establishing common ground. By being perceptually co-present salient events can be made a topic of conversation, making communicative agents aware that they are sharing the same experience.

Designing human-robot dialogue requires an understanding of the consequences of performing tasks that are either related to physical movement or related to managing information. It is also necessary to understand the robot as a unit comprising of both physical action capability *and* communicative capability that together form an integrated interface.

- Dialogue designers approaching human-robot communication need to consider guidelines that are derived from principles for communicative quality, like Grice's Maxims (Grice, 1975), as well as on practical work on non-robotic information-oriented natural language dialogue system.
- In human-robot interaction the communicative functions that has been most in focus concern the tasks: action-directives and communicative acts for specifying long-term goals.
- Communicative functions that change the information context, such as assertions, information requests are also important, as a robot is a part of an information system. Last but not least, meta-communicative acts, like feedback or information on the communicative status, are needed to be able to give help and provide the ability to repair dialogue.
- Grounding strategies need to be selected based on the possible consequences of the robot's actions in the physical and information context of the system. In robotics, cautious grounding strategies can be used when tasks have effects on the physical environment.

CHAPTER 3

Eliciting Human-Robot Communication

When creating a system that does not yet exist, like a service robot, it is necessary to use methods that allow designers to explore and understand interaction design before embarking on large-scale development of working prototypes. In this chapter I will focus on how to use prototypes in the initial phases of system development and how they can be used to elicit data on human-robot communication.

3.1 Filling an experiential void

The creation of an interface that uses natural language as its primary interface model is to some extent a very different undertaking than developing a graphical user interfaces for the standard desktop computer.

Initially both designers and potential users lack mental models of what robots do, and what their interfaces should look like. Most people have no real life experience of interaction with robots. This means that to some extent robot design activities take place in an experiential void. This situation gradually changes as robots become more common in society and people become more experienced. It seems that when faced with robots for the first time humans may have to resort to fictive accounts of *"what robots do"*. Movies, books and comics are filled with friendly[1] robots that speak with creaky voices and understand English perfectly. It also seems that people turn to fictive accounts of human-robot interaction when they try to figure out how to interact with robots. Imagination is an important

[1] somewhat paradoxically, they also typically turn evil and attempt to overpower mankind

and strong force that is essential for the creation of scenarios of future use. However, we have to be sure that claims and requirements made in scenarios can be scientifically motivated. Sometimes there is a thin line between design activities with the goal of developing service robots and philosophical investigations into computational intelligence. For instance, Asimov's "Robot Laws"[2], are discussed by Clarke (1994) as a hypothetical scenario. In relation to design, Norman used Asimov's laws to serve as an inspiration to explore possibilities for human-robot interaction research (Norman, 2005). Although interesting, the step between philosophical investigations and practical robot design appears to be huge.

By allowing potential users to interact with prototype systems the design process becomes a process where new insights about human-robot interaction are gained at each new encounter both for the users as they experience robots in realistic settings and for the design team who can analyse interaction and discover problems and possibilities. In the following I will examine methods that can be used in an explorative manner to collect data from use scenarios to provide a basis for analysing and modelling human-robot communication.

3.2 HRI as a research-driven design process

The development of a communicative interface for a service robot in a multidisciplinary research setting concerns a set of design-oriented activities on several levels (see Figure 3.1):

- *Technical assessment and creation of use scenarios*, where possible robot platforms, sensors, software components and ways of using these to form a complete system for some user group are explored,
- *Interaction design, prototyping and technical development*, which involves the creation of a system that can be tested and evaluated.
- *User studies and analysis of use* are activities that are intended to evaluate the prototypes with a focus on iterative development the technology, the research approach and the interaction design.

[2] Asimov's robot laws define normative ethics for robots in human service (Asimov, 1968).

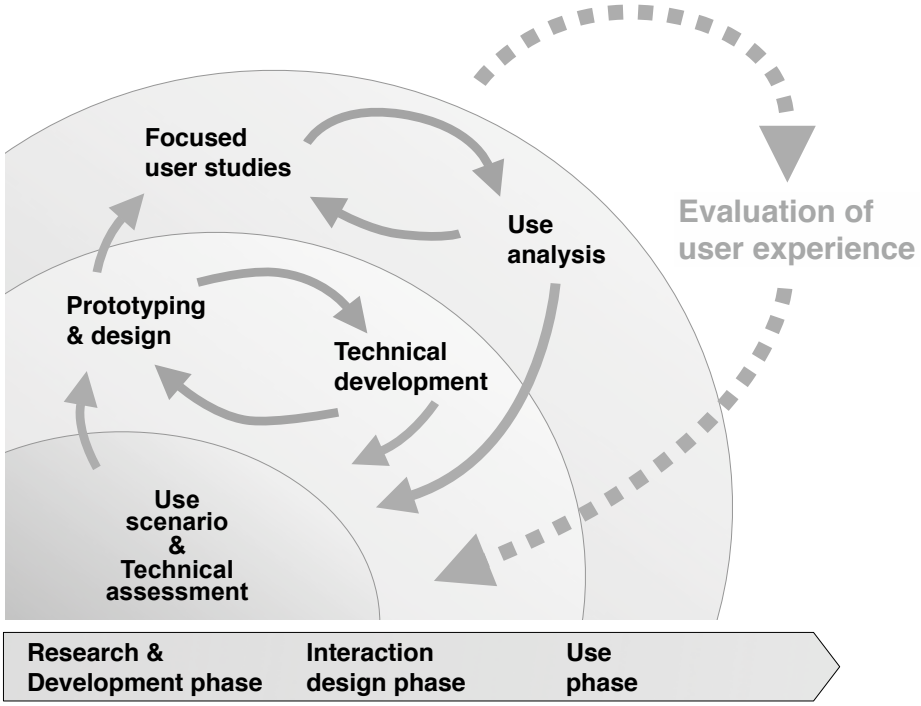


Figure 3.1 *Development of robots with communicative interfaces can be viewed as a research-driven process focused on bringing service robotics to a state where evaluation of user experience on a large scale is possible.*

The sub-levels visualised in Figure 3.1 can all be seen as early steps in a user-centred design process comprising testing with working full-scale prototypes with a usability and user experience focus.

Technical assessment is a strong determinant of the set of possible use scenarios. In a typical research project a technical platform, for instance a research robot platform, is purchased or developed based on general assumptions and considerations of available technology and the type of research problems that will be in focus for the next few years. Once a project has decided on a specific technical platform, the technology that is intended for it, in terms of sensor and actuator capability together with technically oriented research constrains the type of use scenarios that are possible to address from an interaction design point of view. The process of constructing prototypes goes hand in hand with the technical development, providing possibilities and challenges for the design effort. Once a prototype can be

designed and constructed, studies and analysis of use provide new insights that influence the research problems and the technical development. Changes in the base setup, in terms of new sensors or new actuator capability are rare, instead the design loop tends to focus on what is possible without changing the basic technical setup of the system.

The last loop of the model, depicted in Figure 3.1, Evaluation of user experience based on actual use (usually for during a long period of time) is very much dependent on the ability to construct a prototype that is robust enough to allow for real life user testing. When viewing research on communicative robots in terms of a classical industrial view of the product life cycle, where iterative development of prototypes are based on large scale, end-user evaluations, the loop labeled Evaluation of user experience is rarely closed. There are exceptions, but in order to reach this level, the robots being investigated should not be unique single instance research prototypes. To allow for long-term testing with several users in several sites, there has to be a fleet of robots available. This is especially important as users of robots may form communities to get help, assistance or to share experiences (c.f., Kahn et al. 2004, 2005). When robot technology becomes mature enough extensive long-term studies of use can be performed. This is something which has been done for consumer products, like vacuum cleaner robots (Forlizzi and DiSalvo, 2006) and robotic pets (Kahn et al., 2006), and to some extent with research prototypes equipped with natural language interfaces, for instance, Kanda et al (2007) who studied RoboVie interacting with children and Gockley et al (2005) who studied interactions with the roboreceptionist Valerie.

3.3 Use scenarios

Creation of scenarios allows designers and users to form visions of a possible future product. Bannon (1991) phrased this as: "[U]sers need to have the experience of being in the future use situation, or at least an approximation for it, in order to be able to give comments of the advantage or disadvantages of the proposed system.".

Several methods for approaching interaction design on the earliest stages of the design process are available in the field of human-computer interaction. One example which has been used the context of this work are *Focus groups*. The goal was to collect qualitative data to gain an understanding of the attitudes and

background knowledge of users of a future system. A focus group is typically set up as a group interview with 5-10 participants selected based on the characteristics they share. A test leader facilitates and controls the flow of the discussion around a set of relevant questions.

For a focus groups session to be successful there should be some input in the form of a prototype or some concrete ideas communicated by the designers. In order to engage potential users in creative activities there has to be a concrete idea or prototype available (Schrage, 2004). Without this seeding the sessions may appear too abstract to the people involved.

Sketching

For interaction design that involves human-robot communication, ideas and notions of a design team become manifested in a use scenario or a prototype can be seen as a type of agenda-setting. In other words, the point of departure is already decided. Initially it is therefore possible to work with methods that require little or no user involvement. Sketching provides a quick and dynamic way of creating prototypes based on experience (Buxton, 2007) and has been used informally and continuously in the research process described in this thesis. An example of how sketching has been used can be seen in Figure 3.2.

Synthetic dialogues

Another method which has been extensively used for the work presented in this thesis, is *synthetic dialogues*. As a method synthetic dialogues unifies two important determining factors that work as opposing forces: *human capability of engaging in natural language interaction* and *machine perception and understanding* that is possible given the state-of-the-art in natural language technology. Together these factors constrain the possible design space of human-robot communication.

Synthetic dialogues are written pieces of text constructed with the purpose of providing a prediction of the communicative behaviour of both the user and the system to serve as a basis for prototype design. The idea of a synthetic dialogue is that it provides an estimation (or educated guess) of the behaviour of both the system and the user. More specifically:

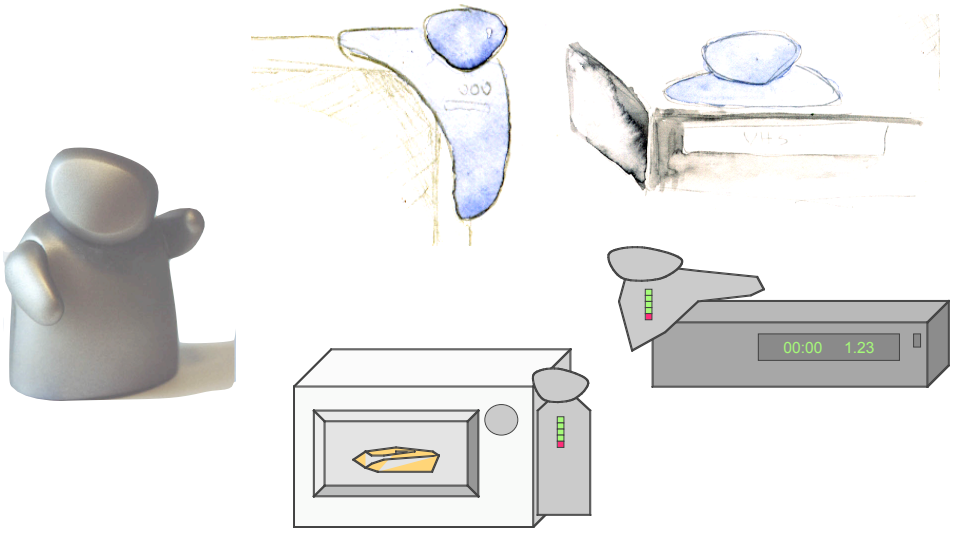


Figure 3.2 *Sketching has been used at various stages of the design process. The images above illustrate some of the sketches that was created to visualise the design of the C-Roids concept presented in (Green, 2001). The original design for the Cero character (left) was created by Erik Espmark and was used as inspiration for concept sketches created to explore the notion of an whole family of interface characters (shown to the right).*

- The dialogue is an honest account, or a vision, of what the system is supposed to handle, meaning that it is "true to the algorithm" (Maulsby et al., 1993).
- The dialogue can be used as an (early) test case during implementation, and for practical purposes serve as a limited *synthetic corpus* in the early design phase.

Variants of the synthetic dialogue method have been used by Torrance (1994) and Isendor (1998) who asked people to write down phrases they believed a robot would be able to handle. It is symptomatic that what they came up with was a set of phrases, rather than dialogues. This is understandable because users were asked to write down what they would say, but were not asked to reflect upon what the robot should say.

From an evaluation point of view, synthetic dialogues can be used to evaluate natural language understanding components during the development process, for instance by running automated tests. Synthetic dialogues should not be the only

measure of system quality. There is a risk that the designer might create dialogues that do not match the dialogues that appear in a real use scenario. Therefore synthetic dialogues should not be part of software requirements, instead they should be used as tools for conceptualisation and evaluation of design in the development process.

3.4 Eliciting communicative behaviour

Real life experience based on encounters with service robots that perform real tasks in real settings is probably an experience that very few people have. This poses a delicate problem when designing a robot prototype. First of all there are expectancies from fictive accounts from movies, TV and science-fiction literature. Second, we have the opportunity of shaping the first encounter with a "real" service robot. A prototype may serve different purpose and be more or less realistic both in terms of appearance and function. The situated character of human-robot interaction makes it necessary to analyse behaviour of people that are encountering robotic artifacts in realistic settings. To create a setting that appears realistic to both users and designers, Hi-fi simulation of interactive behaviours using tele-operated robots can be used. Another approach involves the creation of demo systems that work in a controlled environment, like a research laboratory, a science fair or an office. Both methods can be used to elicit users' reactions, behaviour and attitudes towards service robots.

Wizard-of-Oz simulation

High-fidelity simulation is a methodology used for simulation of high-level functions in an interactive system. The general idea is to simulate those parts of the system that require most effort in terms of development (like a natural language understanding module) or to assess the suitability of the chosen metaphor. In its classical form, where one user interacts with one (desktop) computer in a lab environment, the method has been used for development of communicative systems since the 1970s[3] The starting point for a Wizard-of-Oz study involves the construction of a prototype where some features of the system are provided by real

[3]In the literature good descriptions of how Wizard-of-Oz has been used for natural language development can be found in (Dahlbäck et al., 1993; Maulsby et al., 1993). The term "Wizard-of-Oz" was first used by Kelley (1984) in the eighties.

components and where some functions are simulated by one or more operators controlling the system's actions and responses. A classical setup is to put a user in front of a desktop computer in one room and an operator, a wizard, in another room. The test leader informs the user about the scenario and distributes a set of tasks to solve using the novel system. The interactions between the user and the system are recorded. Since the user often is unacquainted with systems of that particular kind, which is common for speech interfaces, or the task itself, the characteristics of the setup are that of a kind of a role-play, where the user tries to engage and act within the given scenario. The fact that Wizard-of-Oz typically fail to involve users that bring real tasks to the system has been criticised, for instance by Jönsson and Dahlbäck (2000). Allwood and Haglund (1992) have also noted that the wizard operator acting in a scenario is involved in roles on different levels. The researcher role involves acting as a system (wizard operator) and during sessions the wizard can take on different communicative roles like the sender role, the receiver role etc. Bell (2003) noted that in a task scenario, like a travel agency dialogue, the wizard not only acts in a system role but in the role of a travel agent. In reality the behaviour of people acting in a real situation may be quite different from what people do in a simulated scenario even if the user believes that she is interacting with a real system. We might miss out on jargon and lexical phenomena specific to that domain. Distilling real dialogues has been proposed as a way to circumvent this issue (Jönsson and Dahlbäck, 2000). In robot domains, "real" dialogues scarcely exist (for obvious reasons). Nonetheless, the general assumption is that even if people are engaged in a role-play their linguistic behaviour remains consistent even if the conversational partner is a robot.

When spatial aspects and situated interaction become a topic for investigation, the complexity of setting up the scenario increases, especially the amount of people required to maintain and control the scenario. However, when carefully designed, a simulation study will provide data about different aspects of human-robot interaction that would otherwise be inaccessible for analysis until large efforts had been spent on the creation of a working prototype. The data we get from a Wizard-of-Oz study are to a large extent qualitative, but we may also collect data as a resource to be used for component development, for instance, as technical training data for speech recognisers.

A thematic view of the type of data collected using the Wizard-of-Oz technique, that has attracted the interest from researchers yields the following categories:

- Data on language use, especially spatial language (including gesture and speech).
- Enactment of interaction scenarios that appear complete, to allow users and designers to visualise and conceptualise the behaviour of the future system in a realistic setting.
- Assessment of users' attitudes towards a future system or towards robots in general.

Simulation studies have been used in order to investigate hypotheses concerning general aspects of human-robot interaction, such as social behaviours, like studies of spatial positioning (Walters et al., 2005a,b) and collaboration (Hüttenrauch and Severinson Eklundh, 2003). The classical setup of a Wizard-of-Oz study is with one user, but it is also possible to simulate multi-user scenarios, for instance to investigate a robot as a shared resource. A role-play may be used to engage users in a task that lasts for days rather than minutes (Kanto et al., 2003).

Verbal protocols

Another approach to elicit communicative behaviour from potential users of service robots is verbal protocols. The main difference between data collected using a verbal protocol and data collected in an enacted or real use scenario is that verbal protocols are conscious accounts from users that are asked to reflect or comment their actions. In human-robot interaction scenarios the following verbal protocols have been employed in design related activities:

- Written verbalisation
- Post-verbalisation
- Think-aloud protocols

Torrance (1994) used *written verbalisations* as a method to assess users' conception of robot oriented dialogues, or rather dialogues created by introspection about users' own performance. Torrance asked some users to write down what they would say to the robot. He also asked them to rank these sentences in order

of difficulty as perceived by the users. A common approach in usability assessment for development of graphical user interfaces is think-aloud protocols, where the user is told to verbalise his or her actions. In development of speech interfaces this method is hardly ever used, because the obvious conflict in use of the verbal channel.

A related technology, *post-verbalisation*, has been proposed and used by Karsenty (2001). Instead of a continually commenting on his own performance, the user is prompted to comment on the system's performance or to formulate an utterance that he finds appropriate at that point in interaction. This technology has been used in a slightly different variant by James et al (2000). By using pre-recorded scenarios the user was able to view a screen visualisation of the system and the spoken dialogue. During some points (typically in the end) the user was prompted to complete the dialogue, meaning that the user should verbalise what the robot should do next or to rate the dialogue using some evaluation measure. Post-verbalisation is also a way of keeping the design process open-ended and can therefore be seen as co-operative prototyping (Bødker and Grønbäck, 1994). By engaging the user in a language game the evaluator, or preferably the designer, can let the user hear or formulate different responses at some points in the system. In a post-verbalisation session the designer could efficiently act as a representative for the system, providing utterances with certain qualities, while at the same time, keeping limitations of a certain kind of system in focus. To achieve this the designer needs to know what type of actions the system is supposed to be able to perform. Restrictions range from what words that can be used, what type of utterance that the system can parse and what actions the system can be perform at the task level as well as restrictions on conversational behaviours in general. In some respects a post-verbalisation session can be seen as a limited, partial and overt type of Wizard-of-Oz scenario. The important differences are that the user is collaborating with the designer, knowing that the system is being simulated as a low-fi prototype.

For evaluation of interactive systems comprising graphical user interfaces concurrent *think-aloud protocols* (Ericsson and Simon, 1980) have been widely employed, but for obvious reasons not for interfaces comprising a spoken dialogue interface. Another protocol based approach tests the system in retrospect. This is referred to as retrospective think-aloud protocols or retrospective testing (Nielsen, 1994). The idea is that an interaction session is recorded on video. In the fol-

lowing post-session, the evaluator views the session captured on the tape together with the participant while asking the participant to reflect on the interaction. While concurrent think-aloud protocols put constraints on the users' ability to use the voice for actually commanding a robot, it also provides a raised workload, or make users structure their tasks differently. It has been shown that problems found using retrospective testing are verbalised to a higher degree and therefore reveal other types of problems than those found in comparable concurrent think-aloud protocol studies (Van Den Haak et al., 2003).

Synthetic dialogues (see Section 3.3, above) can be seen as a special form of verbal protocol. The main difference is that instead of asking potential users about what they would say to a robot, the designer take both the role of the user and the system at the same time.

3.5 Chapter summary

User-oriented design activities for human-robot communication are performed in what can be characterised an experiential void. This is a state when both users and designers have vague ideas and are uninformed about what a robot can do. As most people have little experience with real life robots, inspiration and models for use may come from robots portrayed in science fiction about how a particular technology like natural language user interfaces can be transferred to a robotics scenario.

Creation of service robots can be viewed as a research-driven process rather than a classical product development cycle. Technical considerations, focused user studies of specific phenomena are necessary before robots can work in real use scenarios. Even if focused studies of user-experience can be performed with research prototypes, assessment of usability and user experience requires robots that are commercially available.

To initially understand human-robot communication it is necessary to create scenarios. This can be done with various methods, such as sketching, synthetic dialogues and verbal protocols. To understand human behaviour in a realistic use scenario it is necessary to create prototypes, for instance by using simulation techniques, like Wizard-of-Oz, that allow designers to enact robot behaviour with simulated and real components working together. To be useful in the design process the simulated user interface components need to be sufficiently constrained with

respect to how natural-like and competent the robot should act within the enacted scenario.

The data collected in a Wizard-of-Oz study range from quantitative and qualitative data on language use to visualisations that allow assessment of attitudes of potential users. The Wizard-of-Oz technique also allows the users to enact the behaviour of a robot – an experience that is invaluable when it comes to the design and implementation of the real system.

CHAPTER 4

Design of Natural Language Communication for Cero

The overall goal for the interdisciplinary Cero project was to create a robot (Figure 4.1) that supports a person with a walking impairment, by assisting with transportation of ordinary objects found in an office or small personal belongings. As this project to some extent could be viewed as a user-centred development of assistive technology, the needs and opinions of the primary user had great influence in the project. Even if the Cero project had a bearing on general issues and challenges in human-robot interaction research, the practical goal of the project was to create a system that could perform practical tasks in the office of a particular user.

The project has also been described by Hüttenrauch et al (2004)[1] who focused on long-term use. The center of attention in this chapter is to give an in-depth description of how the interaction design for human-robot communication was approached in relation to the research questions that were introduced in Chapter 1:

- What is an appropriate communication design for an autonomous robot of this type?
- How can we approach evaluation of the quality of interaction.

To attempt to answer these questions I will provide an account of the findings and observations made in relation to the design and evaluation of the spoken language

[1] See also the technical report (in Swedish) which describes the project in detail (Severinson Eklundh et al., 2001)

Figure 4.1 *The Cero robot with its interface robot and the cover that was fitted on top of the Nomadic Scout.*

user interface. In the end of this chapter I will discuss what we learned from designing and evaluating the communication model for Cero.

Overall goals of the Cero project

The overall challenges and goal of the work with the Cero system was then focused on the development of interfaces that would allow an inexperienced user to specify tasks for the system in an intuitive and "natural" manner (Oestreicher et al., 1999). In the initial stages of the project a questionnaire study and interviews for task analysis were performed (Kahn, 1998) and (Oestreicher, 2002). Based on the findings from these studies we decided to explore two interface models to would

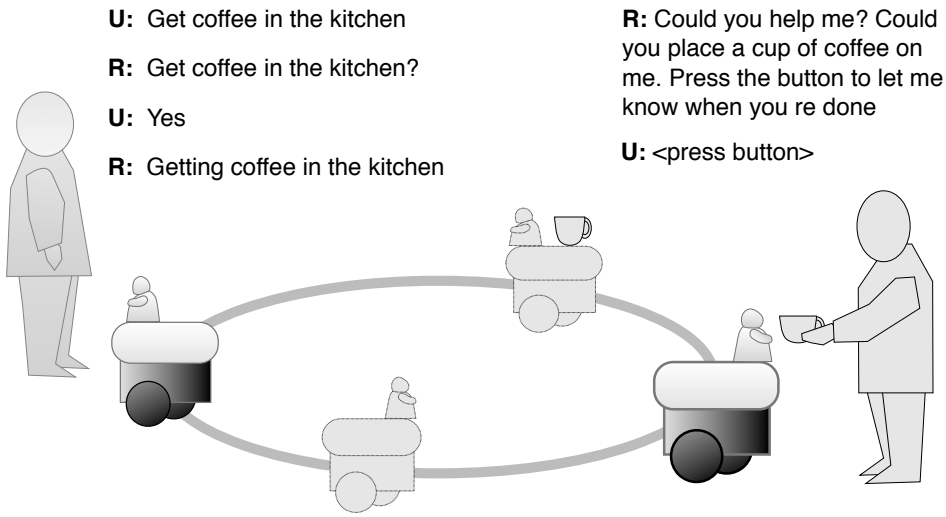


Figure 4.2 *A possible fetch and carry scenario with Cero. The user provides a goal for the robot and the robot navigates to a remote location. At the remote location the robot plays a message and asks for confirmation (someone pushes the button located on the robot). Then the robot navigates to the location where the mission was initiated.*

support the user's work tasks: a graphical user interface and a natural language user interface. A possible instantiation of the system with a natural language user interfaced is depicted in Figure 4.2. From an abstract point of view, these tasks can be seen as two general ways to support the user in her daily activities:

- The robot moves from one place to another based on the initiative of the user, to fetch or to deliver an object.
- The robot carries objects while continuously following the user while he or she is moving within the environment.

Practical limitations

The whole project can be characterised as an exploratory pilot study with the goal of creating a *working robot prototype*. One of the primary goals of the project was to study use and to validate the interface design on a long-term basis in a realistic setting (Hüttenrauch et al., 2004). This goal was very ambitious in the sense that we were facing technical difficulties and limitations that forced us to

not use some of the interfaces, or interface components, in the working prototype. One[2] of the pragmatic considerations was to use the graphical user interface instead of the speech user interface. This decision was reached only after a process of performing serious testing, re-design and re-considerations.

The design process

In Section 3.2 the development of service robots was characterised as a research driven design process. One of the first steps in the process was to find out what technologies were available to support the design, for example, robot platforms, sensor technology and control algorithms. Together with the technical assessment initial investigations were carried out both of potential users (Kahn, 1998; Oestreicher, 2002) and of technological components (Tåqvist, 1999). After acquiring a robot platform we performed an exploratory study using the Wizard-of-Oz technique, where we enacted a possible use situation with an autonomous service robot equipped with a spoken dialogue interface. Because this initial Wizard-of-Oz study was carried out with the goal of collecting unconstrained natural language dialogue it provided a starting point in the development process, rather than an evaluation of a particular design. The findings from the initial study informed design activities of the subsequent development/research process along several dimensions:

- Dialogue design and evaluation of prototypes.
  - Focused Wizard-of-Oz studies of different aspects of dialogue design.
  - Practical tests with the implemented dialogue system.
- Design of different graphical user interfaces hosted on desk-top and hand-held computers.
  - Long-terms user studies.

In the following I will discuss the activities that are relevant for human-robot communication, starting with the initial Wizard-of-Oz study.

[2]The follow-behaviour based on ultrasonic sensors that was developed by Tåqvist (1999) was another component that we decided not to include in the prototype.

4.1 Wizard-of-Oz study I: Unrestricted dialogue

As there are[3] no design guidelines or widely accepted methods for the design and evaluation of human-robot communication for autonomous robots, we were compelled not only to consider the design of human-robot communication but also find ways of studying the interplay between human and robots. Consequently we kept an open-mind to what we would eventually find and saw this as a valuable opportunity to explore human-robot communication in a realistic, but limited, setting.

For the initial study we adapted the classical Wizard-of-Oz method to a robotic setting without constraining the use situation, neither with respect to mimicking the capability of speech recognisers nor with respect to limitations of natural language understanding. In short, in terms of understanding the user, the robot prototype appeared to have a linguistic competence that was well in level with humans. The behaviour of this prototype turned out to be too unrestricted with respect to what could be expected from a real robot in terms of natural language understanding, perception capability and planning.

The robot used in the initial study was a full-scale prototype[4] which was developed based on the Nomadic Scout, which is a robot for robotics research (Figure 4.3). The robot was covered with a casing made out of white foam-core boards with cut out holes for loudspeakers and ventilation (Figure 4.3, right). The robot had a loudspeaker that was used to play synthesized speech. We also devised a simple control system that allowed the robot to be controlled remotely from another room using written commands (Figure 4.3).

One the things that was discovered in the interviews with potential users was that the robot should have some kind of locked compartment where small or valuable objects, like keys, could be placed safely (Severinson Eklundh et al., 2001, p. 18). In the first full-scale prototype a large plastic flower pot, spray-painted in silver, served as a transport compartment. In the later full-scale version of the robot this compartment was integrated with the top cover.

Using this prototype we performed a Wizard-of-Oz study with six participants. We recruited male and female colleagues from the department. They were usability

[3]There are attempts to establish performance metrics autonomous systems (cf. Huang 2007; Huang et al. 2005).

[4]The physical prototypes described in this chapter were developed by the industrial designer Erik Espmark.

Figure 4.3 *The Wizard of Oz setup (left). An early version of a casing for Cero from low-cost materials: a flower pot and cardboard that was cut and assembled to form a simple casing (right).*

experts or programmers, but were not familiar with the technological status of the project nor that we would be using the Wizard-of-Oz method.

The overall purpose of the study was to investigate how participants acted when presented to a service robot and if they showed any systematic patterns during interaction. From the point of view of human-robot communication the purpose of the study was to find out what lexical constructions and what type of phrasal expressions the participants would use. I also wanted to get an impression of what type of dialogue patterns would emerge in the scenario.

We gave the participants a set of tasks to perform using the robot. First the participant should instruct the robot to transport a magazine to another person (which in this task could be considered synonymous with the location of this person's office). Then the participant was to accompany the robot to a table where there was a pitcher with water and glasses available. Using the speech interface on the robot the participant was to instruct the robot to carry the glass of water back to the location that the participants had been told represented their "office" (a table located in the middle of the room).

Two wizard operators controlled the robot. One wizard, the navigator, controlled the movements of the robot. The other wizard, the communicator, controlled what the robot should say by typing messages that were sent to the text-to-speech system. There was no explicit instruction for how the communicator wizard

should handle dialogue, in terms of strategies initiative management and grounding. Due to the constraints in performance, the dialogue was terse, focused on paraphrasing commands (like "I am going to K's office") or acknowledging commands by saying "OK". There was also no instruction for how the coordination between the navigator and communicator wizard should be handled. This sometimes led to situations where the robot started to move before a command had been acknowledged verbally through speech output.

The data from the user study consisted of video recordings from the sessions and post-session interviews. We analysed the data from the study looking for patterns of use and linguistic phenomena that we considered to have an impact on the further design. An example transcription of a user session is shown below:

(27) **U:** OK
(28) **U:** OK
(29) **U:** robot deliver this to K M at room... ⟨pause⟩
(30) **U:** ...1628
(31) **U:** I can ⟨pause⟩ walk with you.
Stands up. Looking at R.
(32) **U:** Are you ready?
(33) **R:** I am going to K
Lets R. pass while observing.
(34) **U:** ahem, follow me, please.
Turning upper body...
...head away from R. walking in direction of K.'s office.
(35) **R:** OK
Looks to K's office. Turns to R
Looks repeatedly back to R.
(36) **R:** OK,
(37) **R:** I am in K's room
(38) **U:** You are not!
⟨ *laughs* ⟩
(40) **U:** Please, two meters left + ⟨DEICTIC-GESTURE⟩

Example 4.1.1: Transcription of wizard data from the exploratory wizard made in the Cero project.

Results from the study

We have not performed an in-depth analysis of the communicative patterns of the dialogue since large portions of the dialogues consists of utterances that can be interpreted as monologues rather than a dialogue between a robot and a participant. This has to do with the performance challenges we experienced when we were producing spoken prompts on-board the robot. The wizard operator was typing in text and the on-board text-to-speech system that was used to render sound from the typed prompts was slow (even short words that could be typed quickly resulted in delays of several seconds). The slow response times made the interaction appear as unbalanced, as participants continued to speak to the robot, and in effect stacking commands. They also commented on their own actions and actions from the robot.

Besides from providing means of collecting data in the form of video recordings, the Wizard-of-Oz study gave my colleagues and me an opportunity to enact the future system together with human users. This has also been noted by Maulsby (1993) who noted that the training and knowledge designers receive from acting as the system is invaluable for informing new design.

When we watched the videos from the user study we observed the following:

- Participants lacked a sense of where the robot was heading as it cylindrical, making it hard to predict in which direction it was moving.
- The robot provided little, or late, feedback through speech. Also physical actions, which could be interpreted as evidence of understanding, were performed slowly.
- Participants closely monitored the physical actions of the robot, taking even small movements as an account of acknowledging participant's input.
- Although the participants in the study had been told that the robot neither was able to detect a participant, nor could interpret gestures as input, several participants used gestures to accompany their spoken actions.

Implications for design

In terms of dialogue design the impressions from the initial study was that design principles and methods for other types of natural language user interface could not be applied directly to interfaces involving human-robot communication. When the first study was performed, the project focused on problems that in a sense can be viewed as a kind of heritage from the project's initial focus on research on natural language understanding. This meant that the goal of the initial study was to establish what people would say to the robot and how these expressions could be interpreted in some type of semantic framework. This is well exemplified by this quote from a position paper concerning the project, written some months before the first[5] Wizard-of-Oz study. The focus is on challenges related to interpretation, ambiguity and issuing of clarification dialogue rather than on challenges related to giving feedback and handling multi-modal input: "The mobile robot and the user are physically in the same room. The robot is told to 'go left' – *dependent upon the location of the robot in regard to the user, the 'correct' execution might mean two different directions. This will need to be resolved by the robot detecting this ambiguity (and solving it intelligently) and/or initiating an appropriate dialogue.*" (Oestreicher et al., 1999, italicised by the author)

After having experienced human-robot interaction with the kind of realism provided by a Wizard-of-Oz study, it became clear that there were other challenges that would have to be addressed alongside problems related to natural language understanding. By reflecting on the observations of the first study, these challenges were formulated as design goals:

- The dialogue system should be designed in such a way that it could engage its underlying planner to perform service tasks in a transparent and reliable way.
- Provide appropriate feedback, meaning feedback that is timely and relevant given the task.
- Provide a design that supports learnability to facilitate for first-time users.

As we also knew that the robot we were designing would also be equipped a graphical user interface hosted on a separate desktop computer. In the duration

[5] See Chapter 4

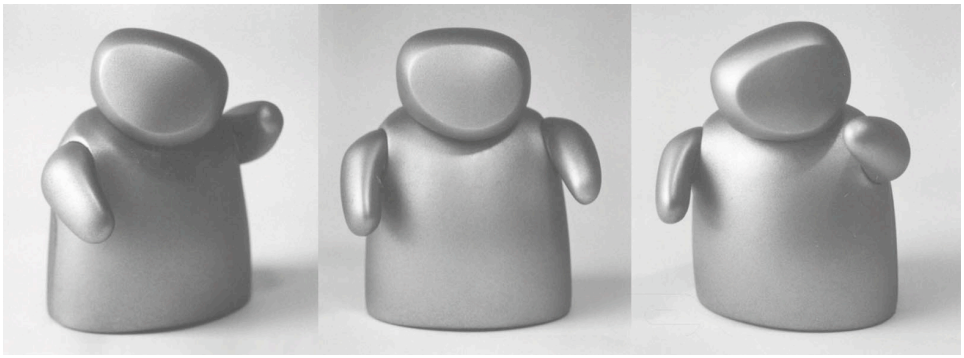


Figure 4.4 *The Cero interface robot.*

of the project we developed a graphical user interface and a speech interface in parallel and one of the challenges we faced was how to design the system so that it would allow a transparent an intuitive transfer of the interaction between the spoken language user interface and the graphical user interface.

4.2 System architecture and services

This section describes the Cero system as a design concept based on services, interface functionalities and software modules that either could be described as working components, or as non-working prototypes where the functionalities could not be made robust enough with the resources available and in the time-frame of the project.

System functionality and architecture

To provide a more complete picture of the Cero project we will briefly describe the system architecture and the basic functionalities of the Cero system. This description is based on work presented in (Severinson Eklundh et al., 2001). The robot was built on the Nomadic Scout, a wheel-based robotics research robot. The robot was equipped with a 266 MHz Linux PC which hosted the control system.

The interface robot, named Cero, was attached to the cover of the transport robot. The interface robot provided an anthropomorphic focal point for the whole robot (see Figure 4.4). We have developed a set of gestures that the interface robot could display during interaction. For example, it could nod to give conversational feedback, and move its arms as if it was walking, showing that the robot is on

its way. The on-board computer was connected to the interface robot trough a serial interface to a microprocessor (a BasicStamp II). The interface robot had four servo motors which, together with the movements of the robot platform, gave it five degrees of freedom. The communicative gestures displayed by the robot are described in Section 4.3 and in Figure 4.8. The interface robot was intended to provide affordances related to communicative behaviour by encouraging the use of human communication patterns. The same kind of character could be used in interfaces of other applications than robots (cf. Green 2001).

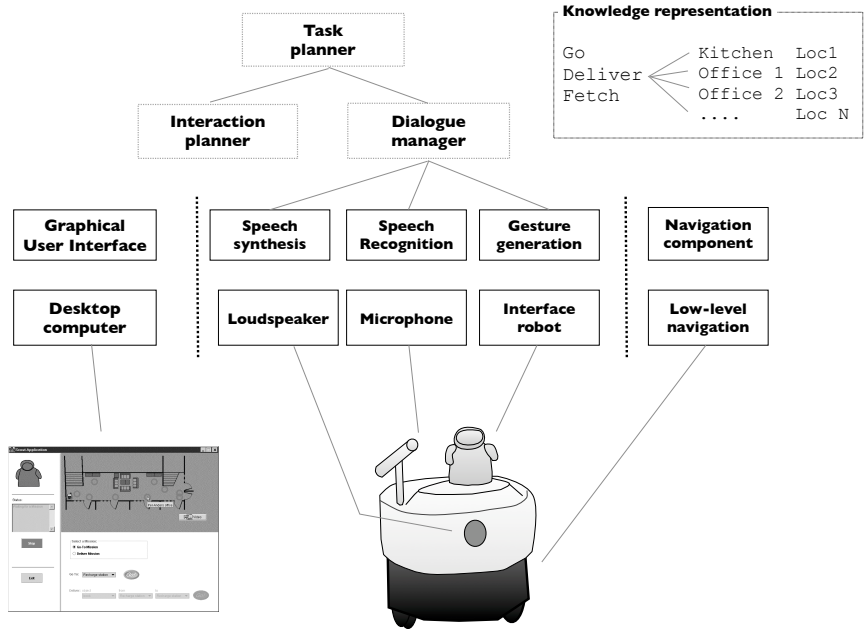


Figure 4.5 *A conceptual sketch of the different layers of components of the Cero system.*

The Cero system consisted of a set of modules that managed the different services in the system. The communication between the different modules was handled using inter-process communication. A conceptual view of the system is shown in Figure 4.5. The core functionality of the system was realised using the following capabilities:

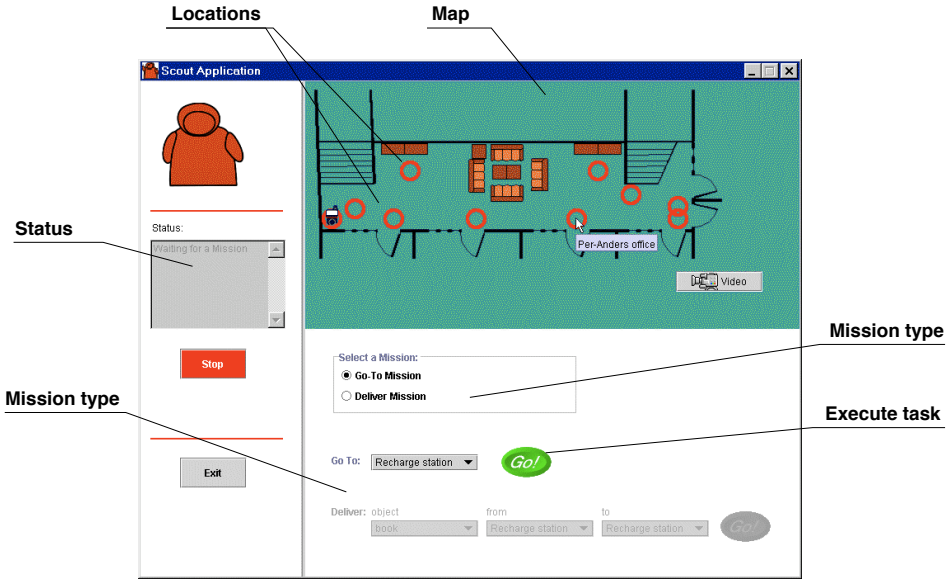


Figure 4.6 *The graphical user interface for the Cero system. The map shows locations that the system knows and can go to as (red) rings. The mission type makes it possible to select whether an object should be carried (delivered) to a location or if the robot simply should go to a location (go to). The known locations and the objects available are shown as drop-down menus on the bottom of the user interface. The green button with the text "Go!" tells the robot to execute the specified mission. To the left status messages are shown during the execution of the missions.*

- Reception of goal specifications by use of speech input or input from a graphical user interface.
- Speech output: the ability to play synthesized speech or audio messages.
- Navigation: the capability to move around in the environment while remaining located in relation to a predefined map.

Services

The tasks the robot could solve using its represented domain knowledge, locations, object, routes between locations and direct movements, were the following:

- *Move to a location* — to move to a location the robot uses its navigation system to drive to a named location described by coordinates in the predefined map.
- *Receive objects* — corresponds to the act of a user placing an object in the transport basket of the robot. The robot has no ability to handle objects through the use of a manipulator, nor does it sense that this act has been performed.
- *Deliver objects* — to deliver objects the robot needs to represent a carried object and a goal location. Once at the location it may play a verbal utterance (or another sound signal) that it has arrived at the goal with a specific object.
- *Direct movement* — the robot could handle navigation commands that circumvented the use of the location-based navigation system. These commands allowed the robot to move forward, backward, and turn.

The task-planner connects to the interaction manager which handles the communication with the user interface (the GUI or the speech interface) and controls when the robot plays synthesized speech while away on a mission. The task-planner uses the navigation component to move the robot and receives messages once the robot has completed a route.

Apart from playing synthesized speech the interaction manager also controls the commands that are sent to the interface robot placed on top of the robot cover.

To navigate the robot uses an implementation of SLAM[6] (Andersson et al., 1999; Wijk and Christensen, 1999) which uses sonar landmarks from the 16 ultrasonic sensors and information about the distance travelled from the odometer as input to keep the robot localised in the environment. The sensors are also used to avoid obstacles, objects that are not found in the map. Maps have to be predefined so that landmarks can be recorded manually by driving the robot around in the environment. Once landmarks have been recorded in relation to the map, the localisation and navigation modules of the system provide the services needed to allow for navigation commands to be handled by the system.

[6] Simultaneous Localisation and Mapping

Task-related goal types

The type of actions that the system can perform in order to solve the tasks are physical movements, communicative acts or virtual acts (which involve changes in the internal representations of the robot). Using the categorisation from Section 2.5 the task-related goals are the following:

Protractive goals: concern actions where the robot executes several actions as part of a plan or script. These goals should be handled as continuous activities that are stretched out both over time and in the physical environment. Missions are lasting minutes and hours, rather than seconds. The distance covered is in the magnitude of hundreds and tens of metres rather than centimeters.

Directive goals: concern actions that are handled as single instances concerning short distance and short time spans. Missions lasts seconds rather than minutes. Distances range from millimetres to centimetres.

Information-oriented goals: concern actions that are solved using information processing and communication with other systems. These goals are handled without engaging physical actions of the system.

The specific goals that the system needs to support are listed below, together with examples of possible natural language input:

Protractive goals	
GO-TO	Going to a location starting in the current location
Example:	"go to K's office"
GET	Get an object from a location,
Example:	"get a cup of coffee from K's office"
DELIVER	Deliver an object to a location
Example:	"deliver a cup of coffee to K's office"
REQ-ASS	Get assistance from a secondary user
Example:	by saying "Can you place a cup of coffee on my tray?"

Directive goals	
MOVE	Move to a specified direction
Example:	"robot move left"
TURN	Turn in a specified direction
Example:	"turn right"
STOP	Stop the movement
Example:	"stop robot"

Information-oriented goals	
INFORM-STATE	Inform the user of the current state
Example:	Answering the question "what are you doing?", (inform about the current mission).
CHANGE-STATE	Change a system state
Example:	Turn the sound on or off, by saying "silence".

4.3 Dialogue design for Cero

The exploratory Wizard-of-Oz study, described in Section 4.1, provided useful insights into the challenges of design of human robot communication. Based on the observations from the exploratory study and a thorough assessment of the technology a dialogue design was created. The goal of this design is to handle scenarios like the one depicted in Figure 4.2 where the user sends the robot out on a mission that involves goal specification and involvement of other people as helpers. To handle dialogue for Cero we cannot solely focus on dialogue that is task-related. To achieve a system that is perceived as responsive and attentive we also need to consider various ways of providing feedback. The dialogue design for Cero focused on these aspects:

- A dialogue strategy based on grounding.
- Task-specification of long-term (protractive) goals.
- Task-specification of short-term goals (directive) goals.
- A feedback model that included the use of an embodied character.

I decided that although it was possible in principle to handle information-oriented goals using the system, this would not be investigated in close detail as it was out-of-scope with respect to the overall project goals.

In the following sections we will describe the dialogue design starting with the overall dialogue handling strategy.

Dialogue handling strategy

In the work with the spoken dialogue interface for Cero, we have focused on providing a dialogue that takes human dialogue strategies into account. Rather than using a strictly command-oriented approach, where a natural language command is reactively followed by an action, we have developed a dialogue model based on the principles of grounding in human-to-human dialogue (see Section 2.3).

The tasks that are accommodated in the system concern going to places (for instance, "go to Mary's office"), fetching and delivering object ("Get coffee from the kitchen", "Deliver coffee to Mary"). To solve these tasks, proper phrasing of the system's contributions is an important part of the process of achieving common ground in dialogue between the user and the robot.

We intended that the system would use the following strategies to handle the dialogue:

- Get initiative and maintain the initiative.
- Ground dialogue through feedback.
- Error-handling by backing off.

Grounding strategy

At the level of task-specification, we use a cautious grounding strategy (Larsson, 2002, p. 97) to assure that the user becomes certain about what instructions the robot has received and is about to carry out. This means that the robot acknowledges the user's request by reformulating it as a question, requesting confirmation by the user. This may in turn be confirmed, as requested by the system. The robot's request for confirmation may also be rejected. This means that the robot does not start the mission.

In the following example, the robot receives a command that is only partially understood by the dialogue system. The following turn by the user provides more information for the system. If the user responds by specifying a location in the next turn, the system may infer that the task of getting and the object to be collected are part of common ground:

(41)	**U:**	Robot, get a paper	⟨Request paper⟩
(42)	**R:**	Get a paper from where?	⟨Request location⟩
(43)	**U:**	From John's office	⟨Specify paper⟩
(44)	**R:**	Get a paper from John's office?	⟨Request confirmation⟩
(45)	**U:**	Yes	⟨Acknowledge⟩
(46)	**R:**	Getting a paper from John's office!	⟨Report action⟩

The directive instructions used for near-navigation, for instance"turn left", are not grounded to the same degree as the fetch-and-carry tasks. Here, an optimistic strategy is used so that only the first directive command given by the user is grounded by asking for confirmation. If a new directive command is issued, it is carried out reactively, assuming that the user is in close vicinity of the robot and is monitoring its movements:

(47) **U:** Move forward

(48) **R:** Moving forward

(49) **R:** ⟨moves⟩

The system provides explicit feedback on received commands, and makes clarification requests if there is some information missing according to the domain-oriented consistency check. This assures that only fully specified planning constructs are sent to the system's planner. A cautious grounding strategy is used, requiring confirmation from the user before actually attempting to perform any physical action with the robot.

To accommodate directive interaction that will allow for specification of short-term goals for near-navigation, an optimistic grounding strategy is used. Once a command such as "move forward" is received these commands are assumed to be grounded immediately once they have been received by the system.

There is one exception from the optimistic strategy, and that is when there is a shift between "modes", meaning when the robot receives a directive command, for instance "move forward", just after having performed long-term command, for instance "go to the kitchen", a pessimistic grounding criterion is used, like "move forward, yes?".

(50) **U:** Get paper
(51) **R:** Where is the coffee
(52) **U:** No!
(53) **U:** What is the object?

Example 4.3.1: An example of the handling of miscommunication in the Cero dialogue system

The error handling in the Cero dialogue system should contribute to the system being perceived as flexible and intuitive in a use situation. In Example 4.3.1 the speech recognition component miscrecognises the utterance *"Get paper"* and instead perceives *"Get coffee"*. When the system tries to ground this task, by providing explicit feedback, asking for the missing location, the user may recognise that the object is wrong and reject the task by saying *"No"*. The system then backs off by assuming that the overall task (*"get"*) is correct and tries to fill the first frame by asking for an object.

Supporting specification of long-term goals

In the Cero project we used synthetic dialogues for different purposes. The synthetic dialogue that was developed for Cero primarily concerned predictions, based on experience and intuition of how users specify long-range and directive goals.

The long-range goals that could be specified by the user in the Cero system concerned the services GO (to a location), DELIVER (something to somewhere) and FETCH (something from somewhere). As we were interested in providing dialogue capability for handling grounding we constructed dialogues containing the phenomena we judged that the system would handle:

(54)	**U:**	Robot, get coffee	Partial: no location
(55)	**R:**	Get coffee, where?	Request location
(56)	**U:**	In the kitchen	Specify location
(57)	**R:**	Get coffee in the kitchen?	Request confirmation
(58)	**U:**	Yes	Confirmation

Enabling specification of directive goals

The term directive dialogue concerns dialogue which is used to instantly controlling the physical behaviour of the robot. When a user is focused on controlling the movements of the robot directly, in the same manner as with a joystick, the features of the spoken language modality means that the user has to resort to guiding the robot using directive commands.

The specific robot platform that is used will affect the way directive dialogue can be carried out. When this is operational the behaviour of the robot as it strives towards a goal while avoiding obstacles gives it a characteristic behaviour: as the robot approaches a goal, it is continually making adjustments in the heading while it slows down and finally stops. At this stage the position may not be optimal with respect to what the user is trying to achieve or what the user is expecting. The robot may be too far away from the goal point or it might be heading the wrong way. This makes it hard for the user to load objects and the need for making corrections becomes clear.

There are other characteristics of the robot that may affect the possibilities and capacity of the users to control the robot using directive commands:

- The robot cannot move backwards, because the help wheel placed on the back of the robot constrains movement.
- The robot turns on the spot, steering is handled by individually controlling the speed of the two main wheels.
- The robot cannot move sideways. The robot has no omni-directional drive.

Directive dialogue involves goal specification for short term, immediate action. System goals are then completed near the user and during a short time span. Another characteristic of directive dialogue concerns how "aware" the system is of the user's goals and plans. When the user has specified long-range goals, the robot keeps these in memory while carrying out actions that form part of a (long-range) plan associated with a particular goal. This means that the robot may take the initiative in dialogue, as a part of a plan, for instance by asking someone to put an object on the transport tray. When the robot executes directive goals there are no long range plans associated with these goals. Instead the initiative is on the part of the user, and actions are performed as part of the user's plan. This means that

the robot is self unaware of any higher goal the user might have. Consequently the robot cannot prioritise between different goals, meaning that the human performs all the planning. This also means that the robot may perform actions that have dangerous consequences, for instance, telling the robot to "go forward" when facing a staircase might mean that the robot crashes down.

The following set of movement capabilities available on the system were used as a starting point for the dialogue design:

- Forward movement
- Left/right turn
- Stop

Using these movement capabilities I constructed synthetic dialogues for directive commands. During this process some challenges related to design emerged. One immediate challenge has to do with turning left or right. The problem with this command is that it is ambiguous relative to the robot's point of view (RPOV) and user's point of view (UPOV). We assumed that the physical act of turning the robot would disambiguate the dialogue, by displaying an action that would manifest that the robot used an intrinsic point of view:

(59) **U:** Turn left (UPOV)
⟨ *robot turns left RPOV* ⟩
(61) **U:** No, the other way!
⟨ *robot turns left RPOV* ⟩

In this dialogue the situation coincides with the user assuming a robot-centric point of view, something which does not lead to a subsequent contribution from the user with the purpose of repairing the dialogue:

(63) **U:** Turn left
⟨ *RPOV* ⟩
⟨ *robot turns left RPOV* ⟩

It seems reasonable that the user will be able to shift to mean "left" with respect to RPOV in the dialogue that follows. In a context where the task of turning the wrong way has serious consequences it would be necessary to assume a cautious grounding strategy similar to the what has been discussed in previous[7] sections.

[7] See Sections 4.3 in this chapter, and 2.5, p.2.5 in Chapter 2

The the way the navigation is performed by the robot will affect the way the dialogue for move commands needs to be designed. In the case of a platform that could make a direct move towards the specified direction (without turning around) the issuing of a move command could cause the robot to move directly towards this direction. In this case the platform would still be oriented towards the user.

Since the robot platform cannot perform direct sideways or backwards movement it was necessary to consider what turning the platform might cause in terms of changes in view-point.

- *Driving to a point located to the left/right of the robot*. By navigating directly to a goal point located to the left/right of the robot, it will start by driving forward before turning to the left or right. This has to do with the way the navigation system works and might cause problems if the user is very quick in her reactions and tries to change or revoke the command. A very likely situation when this may occur is when the robot is facing a wall or an obstacle. A reasonable reaction when the robot moves forward would be to say "stop" if the user felt that the robot seemed to be going into the wall.
- *Moving backwards* is another case which appears to have similarities with the cases of move left/right since the robot cannot move backwards directly, but has to turn around first. If the robot could back directly, it seems possible that the movement of going backwards would be very much like forwards in terms of expected dialogue structures. When the robot cannot back directly we have to turn a full 180 degrees before moving backwards. Then, as in the example for move left/right there is the problem of deciding the final configuration.

Until now I have discussed what can be regarded as clear cut cases of the user specifying a task. It is possible that there might be cases which can be considered as follow-up requests, like the following:

- Repeat and possibly modify its last action: *"do that again"*, *"turn a little more"*.
- Totally or partially undo the effects of its last action: *"go back"*, *"no, the other way round" "not so much"*.

There were very few instances of this type of behaviour in the initial wizard-of-oz study. Example 4.1.1, which was introduced earlier, contains tendencies that could be interpreted as follow-up requests with the purpose of modifying the robot's behaviour:

(66) **R:** OK,
(67) **R:** I am in K's room
(68) **U:** You are not!
⟨ *laughs* ⟩
(70) **U:** Please, two meters left + ⟨DEICTIC-GESTURE⟩

Based on these data, it was hard to propose a synthetic dialogue for these aspects of dialogue. Synthetic dialogues is a method which relies on the use of knowledge and judgement of what type of dialogue is possible and likely to occur. From this point of view this represented a boundary for the design at this stage. Taking design to the next stage in the development cycle requires testing and practical evaluation.

The feedback model of the Cero system

Users in the exploratory Wizard-of-Oz study seemed to experience a general lack of feedback when interacting with the system. This soon became an important focus in the design activities and was addressed by providing communicative feedback along three dimensions:

- Task-related feedback with respect to the domain:
  - Explicit (verbatim) feedback on given input, for instance by repeating that the object and location.
  - Paraphrasing input. When the system has enough information it paraphrases the specified command using the given object and location to generate a paraphrase of the information available to the system.
  - Requiring explicit confirmation: by paraphrasing the input and then require the use to respond by saying "yes" or "no" we acquire explicit permission to perform the task. In this manner the user can avoid sending the robot off on a long-time, long-distance mission without the possibility to interrupt it.

- By providing feedback pertaining to the communicative aspects of interaction such as the perceptual status of the system by:
  - Displaying gestures with the interface robot Cero (see Figure 4.8).
  - Playing a short sound signal, to immediately give positive or negative feedback based on recognition results.

The interface robot had the twofold purpose of providing a clear cue of where the robot is heading for the users, and to provide low-level feedback as a supplement to the spoken feedback issued by the dialogue system.

Feedback level	**State**	**Gesture**	**Prompts**	**Polarity**
0. Not attending	Robot off	⟨no movement⟩	⟨no sound⟩	
1. Attending	Microphone sound detected	RAISE HEAD	–	+
2. Hearing	Speech detected	RAISE HEAD	–	+
	ASR Failure	SHAKE HEAD	Negative feedback	–
	Partial input	NOD	–	+
3-4. Parsing, interpreting	Complete phrase	NOD	Positive feedback	+
	Inconsistent command	SHAKE HEAD	"Specify X!"	–
	Parsing inconsistencies	HAKE HEAD	"Please repeat!"	–
5. Intending	Planner failure	SHAKE HEAD	"Cannot do X!"	–
6. Acting	Execution of tasks	NOD	"Go to X?"	+
7. Reporting	Task performing	WALK	"Going to X!"	+

Figure 4.7 *Feedback levels according to Brennan and Hulteen, and examples of their realisations in the Cero robot system. The category Polarity tells whether the feedback given is negative of positive. The gestures corresponds to the gestures depicted in Figure 4.8.*

The feedback design for the system has been inspired by the categorisation of feedback used by Brennan and Hulteen (1995). They proposed a taxonomy of eight categories or levels of feedback which correspond to different activity levels. Fig 4.7 shows how these levels correspond to system states, with feedback on the various levels given through gestures and speech.

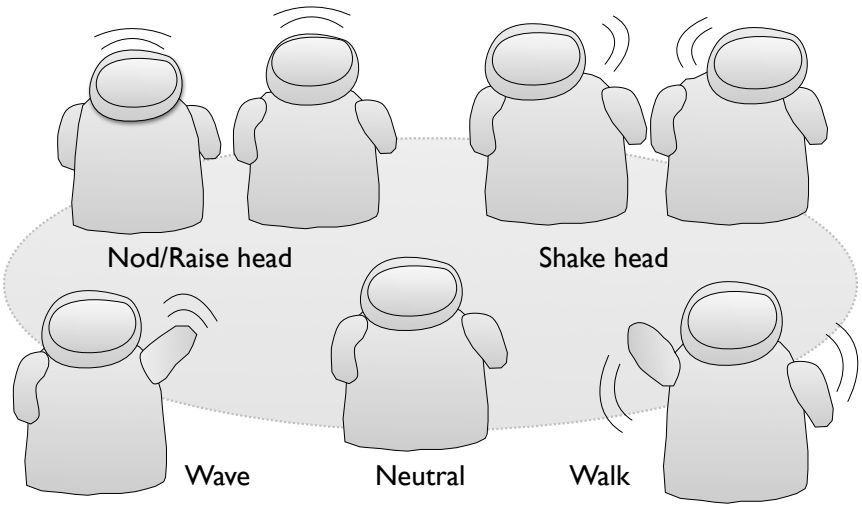


Figure 4.8 *The gesture repertoire of Cero. The movements are designed to be integrated with the speech system so that it is both capable of issuing conversational gestures: raise or lower its head re-actively, based on system states, and co-expressive conventional gestures, like emblems: nod or shake its head, call for user attention.*

Feedback on the first level, corresponding to Level 0 in Brennan and Hulteen (1995) displayed in Figure 4.7, concerns whether the system is active or not. The next level, corresponding to Level 1 in Brennan & Hulteen (1995), concerns if the user has the system's attention, or the system has detected a human voice.

If the speech recognition system provides incremental recognition of commands, the next feedback level concerns partial results issued by the system. This level is corresponding to Level 2 in Brennan & Hulteen (1995). Partial hypotheses that stem from incremental parsing can be expressed by small head nods issued by the Cero character.

Feedback corresponding to Level 3–4 in Brennan & Hulteen (1995) concerns natural language processing and may produce different types of responses of the system. For instance a low confidence score might trigger a request for the user to adjust the microphone, or asking the user to repeat the command. Once the speech recogniser reports that a phrase has been recognised a positive feedback sound is played. If the speech recogniser fails, a negative feedback signal is played. The sound signal for giving positive feedback was a short chirping sound with a raising pitch. To give negative feedback a sound with a falling pitch was used.

Feedback on the task level, which is corresponding to Level 5 in Brennan & Hulteen (1995) provides information regarding the possibility and attitudes towards carrying out the task, for instance by asking the user to confirm an action.

Feedback concerning actions, corresponding to Level 6 in Brennan & Hulteen (1995), is typically signalled by the movement of the robot itself.

Feedback on the last level, corresponding to Level 7 in Brennan & Hulteen (1995) concerns reporting on the current actions of the system (like "Going to the kitchen").

When designing the feedback system of Cero's speech interface, we have not attempted to construct gestures corresponding to all eight levels. Rather, the animated Cero figure is intended to give feedback for which the speech synthesis is not adequate or sufficient, for instance by displaying that speech has been detected or showing that the system is switched on.

4.4 Practical evaluation of the natural language-based prototype

During the Cero project we tested the components that would have to be integrated in order to create a working prototype that used natural language as its main interface modality. With the notion of *working prototype* we mean something that could be put in the hands of test persons with little or no training to allow for use in the fetch-and-carry task that constituted the core scenario of the project.

During the project we carried out design and implementation of different components for the system as if they would be incorporated in the final prototype. The goal was to test the system with users in a realistic situation in a long term study. When testing the components the technical constraints of using speech recognition in a realistic situation became clear (see the following section) and we decided not to use the spoken language user interface in the prototype that was used in the long-term study reported by Hüttenrauch et al (2004). The painstaking work at arriving at this conclusion, which included several practical tests, has still led to some useful observations, which we will share in the following. The practically oriented test sessions of different aspects of the system were performed in the spirit of discount usability methods (c.f. Nielsen 1994) which are focused on the discovery of problems and challenges in the design.

Individual test session with the primary user

This section describes a test done on a version of the system which was installed on a laptop computer, without the robot being physically present. The test focused on aspects of the verbal dialogue and did not include the use of the Cero interface robot.

The dialogue system that was tested using a commercially available speech recogniser for Swedish and a headset microphone. The goal was to give the primary user an impression of what interaction with the system would be like. In this session, which lasted about 35 minutes, the user was proactive and proposed features that were not implemented, already before trying out the system:

- A function for providing help using the phrase: “what can I say”.
- A visual description of the system.

Initially it was observed that the user did not seem appreciate wearing the head-set microphone, even if this was not communicated verbally during the session.

Once the system had been started the interactions did not go smoothly, and the system did not recognise the users spoken utterances. As the system did not respond immediately, the user addressed the test leader and continuously provided meta-comments. This affected the performance negatively by triggering false recognitions. This in turn made the interaction even more confusing for the user.

The performance of the speech recognition was very poor and consequently the degree of miscommunication was high. In the following example, the dialogue almost leads to a state of complete breakdown. Even if the system responded to the input, it did not respond with something that actually brought the interaction forward from a task-oriented point of view. Instead the system persistently failed to recognise any of the user’s utterances:

(71) **U:** Move one meter forward

gå en meter framåt

(72) **S:** Repeat the utterance

upprepa yttrandet

(73) **U:** Move forward

gå framåt

(74) **S:** repeat the utterance

upprepa yttrandet

⟨ *later* ⟩

(76) **U:** Go right

Gå till höger

(77) **S:** Repeat the utterance

upprepa yttrandet

⟨ *To test leader: "No, it does not respond"* ⟩

Till testledaren: Nä, den verkar ha hängt sig

(79) **U:** Go forward

Gå framåt

(80) **S:** Repeat the utterance

upprepa yttrandet

⟨ *Session ends* ⟩

Example 4.4.1: A session with the primary user. The example is originally in Swedish.

These were the challenges discovered in the practical session with the primary user:

- Having no robot present during the tests creates to a non-realistic use situation. There was no scenario or task-instruction that would have provided a background context to what could be said. With no robot present the user does not know what to expect in terms of robot actions.
- The dialogue style was perceived as too strict and unnatural.
  - Questions uttered by the speech synthesis had no question intonation.
  - The terse prompt style was perceived by the user as if the dialogue style should be to use "military style" commands.

- There was a general lack of relevant spoken and visual feedback to signal success and failure in the speech recogniser.
- Wearing a head-set microphone was uncomfortable for the user.

The session with the primary user provided a lot of valuable feedback on how the system worked. From a usability point of view this feedback was invaluable, although the test was much less successful than expected from a technical an practical point of view.

Practical evaluation of task-oriented dialogue

During another practically oriented test session using the task-oriented version of the dialogue system we recorded interactions with a trained expert user, a colleague participating in the project. By evaluating the interactions recorded in the test session I wanted to evaluate how the system handled dialogue specification of the long-range goals: GO-TO, GET and DELIVER. The goal of the analysis was to:

- Evaluate the overall dialogue functionality of the system.
- Analyse the interaction that was recorded on video tape looking for:
  - Behavioural patterns related to interaction
  - Instances of miscommunication.

The user in this study wore a head-set microphone and the speech input was processed on a dedicated computer connected to the robot via the wireless network. We then recorded several sessions where the trained user tried to use the system, adapting his communicative style to how it had been designed. We recorded the sessions and used them as data for an analysis of communication patterns in the system. The types of dialogue that were recorded are shown in Example 4.4.2 and 4.4.3. From the user's point of view, the goal of the interaction was that he should *attempt to make the dialogue flow with as few problems as possible*. The reason for this was that we wanted to encourage a situation when human-robot communication would flow smoothly without errors.

Adaptation during miscommunication

Although the expert user attempted to minimise the cause of errors in the dialogue by talking clearly and in a relaxed tone of voice, miscommunication occurred fre-

quently. One way of overcoming miscommunication on the part of the user was to perform different types of adaptations[8]. In the transcript shown in 4.4.2 the system is slow in its responses, something which is causing miscommunication in different ways.

First of all the system responded slowly to the user's input. This was caused by slow responses from the speech recogniser. Secondly, miscrecognition caused two different types of adaptations to occur:

- The user speaks louder with a changed intonation.
- The user rephrases the command.

Thirdly the sequencing breaks down. The user breaks the sequence which ideally should follow the pattern A-B-A-B, and responds twice after the robot has asked whether it should "Go to Lars office?". In this case the reason for the miscommunication is that the speech recogniser does not manage to translate the utterance to text, something which causes the system to fail at providing a response in time.

(82)	**U:**	Cero!	
(83)	**R:**	Missions: Deliver, Get, Go. Please specify a mission, for instance: Go to Maria's office.	
(84)	**U:**	Go to Lars office!	//5 sec pause//
(85)	**U:**	Go to Lars office!	//5 sec pause//
(86)	**U:**	Cero, go to Lars office!	//2 sec pause//
(87)	**R:**	Go to Lars office?	
(88)	**U:**	Yes	//5 sec pause//
(89)	**U:**	Yes	//2 sec pause//
(90)	**R:**	Going to Lars office!	
		⟨ *robot starts moving* ⟩	

Example 4.4.2: Temporal patterns

I also observed a pattern concerning the time it takes before the user makes another contribution when the system has failed to respond to the user's initial utterance. There are two cases when the response time is important.

[8]For an in-depth study of linguistic adaptations in other contexts, see Bell (2003).

- the time between an utterance and a system response,
- the time between the end of an utterance and the time until the user realises that the system has failed to respond.

It seems that the user in Example 4.4.2 has learned how long it takes before the system issues a command (interpreted by the user as a sign of successful input). In Example 4.4.2 the five-second delay between the first and second issuing of the command "Go to Lars office!" can be divided into a first part and a second part. In the first half of the five-second pause the user first waits a couple of seconds for the system response, then after noticing the failure to respond needs another couple of seconds. The strategy then seems to be to wait at least the amount of time that it usually takes for the system to respond to a command. This may also explain the fact that the user stops speaking in the middle of the confirming utterance "OK" on line 5 ("OK") of Example 4.4.3. After noticing that the robot starts moving there is no need to confirm the action.

		⟨ *robot in navigation state is moving slowly* ⟩	
(93)	**R:**	Put the paper on the tray please!	
		⟨ *user puts the paper on the tray* ⟩	
(95)	**U:**	OK	//7 sec passes//
(96)	**U:**	OK	//overlap with robot moving//
		⟨ *robot starts moving* ⟩	

Example 4.4.3: A dialogue where the user co-operates with the robot at the second part of a fetch mission.

Both the wizard-study and the examples discussed here show that the users closely monitor the behaviour of the system. Users interpret even small movements by the robot as signs of the robot's intentions. In the fetch dialogue (Example 4.4.3, above) the user monitors the behaviour of the robot. It only takes a slight movement of the robot to make the user believe that the robot is about to embark on its mission.

Findings

In summary the analysis of the data from the sessions with the experienced user revealed three types of miscommunication relating to *Sequencing, Response-time* and *Attunement to the system actions*.

Sequencing: The flow of conversation was designed to accommodate contributions that followed the pattern A-B-A-B. There are several accounts where the users did not wait for the contribution of the system but instead made another contribution, something which can be considered to be miscommunication related to sequencing.

It is useful to point out an important difference between the behaviour observed in the initial Wizard-of-Oz study (Section 4.1) and the study reported in this section. In the dialogue system tested in this section the dialogue was designed to follow the pattern A-B-A-B. The dialogues in the initial Wizard-of-Oz study shows another pattern. Instead of issuing a command and then waiting for feedback, the users were stacking several utterances in a row. In the Wizard-of-Oz study we attributed this to lack of timely feedback.

Interestingly, although the feedback in the initial Wizard-of-Oz study (see Section 4.1) was slow, we did not observe strategies or behaviour that seemed primarily related to overcome problems with speech recognition, such as rephrasing or speaking louder. Even if the users stacked utterances, the actions the robot eventually performed seemed to indicate a general understanding of the users' utterances. This is to be contrasted with what the strategy of the experienced user in the practical test session described in this section. The experienced user was well aware of the potentially slow response from the robot, but there was also another circumstance that was important. The experienced user knew that the system was supposed to provide requests for confirmation in order to ground the tasks to be performed, *before* any actions were carried out. There was stacking of utterances in the practical test session, but it appeared in a much more controlled manner, and the utterances were adaptations with similar meaning rather than utterances that contributed new information, as in the wizard-study.

Time to response: Another pattern that was observed was cases when there seemed to be a pattern concerning the timing between the user's utterance and the robot's failure to respond. Two cases appear to be relevant for the problems related to timing:

- Time between an utterance and a system response,
- Time between the end of an utterance and the time when the user realises that the system has failed to respond.

If we attempt to analyse this in terms of turn-taking rules, using the notion of a Transition Relevant Place (TRP) where during a turn there is a component where the current speaker selects the next speaker, "the party so selected has the right and is obliged to take the next turn" (Sacks et al., 1974). In the scenario the utterance "Go to Lars' office" can be viewed as the first part of an adjacency pair *Request-Response*[9]. The preferred answer to a Request is a task-related action together with a confirmation or a rejection of the request. Doing or saying nothing would then be a dis-preferred response (Levinson, 1983) which can be viewed as the current speaker having failed to select the next speaker and therefore the current has to overtake the obligation to respond.

In a more recent study Shiwa et al (Shiwa et al., 2008) used conversational fillers (like the Japanese word *etto*, corresponding to the English *huh*) to reduce the negative impact on the attitudes of users towards long system response times.

Attunement to system actions: We found that the users closely monitor the behaviour of the robot. It only takes a slight movement of the robot to make the user believe that the robot is about to embark on its mission. The last theme that we identified in the practical sessions, the attunement to the actions of the robot has spawned more research questions concerning how spatial features of human-robot communication can be analysed and used in the design. In the practical use sessions we observed that the users appeared to react very quickly to the movement of the robot. From an activity point of view this can be interpreted as *evidence of understanding* (Clark, 1996) and from the perspective of spatial interaction we can see this as *spatial influence*. In this particular context, where the user is awaiting a cue from the robot that means that the robot needs assistance (cf. Example 4.4.3), the attunement of the user was perhaps more related to the robot providing evidence of understanding through its physical actions. In Chapter 8 we will discuss how the robot may influence the spatial behaviour of the user in a comparable, but slightly different way, through the design of *spatial prompting*.

4.5 Wizard-of-Oz study II: Directive interaction

The goal of the design activities concerning directive dialogue introduced in Section 4.3 was to create a dialogue that would handle commands related to direct

[9]The Response part could be "Acknowledge/Confirm" or "Reject".

movements. In this section I will describe a design to handle the specification of directive goals which incorporates the use of a small set of pre-defined gestures. To evaluate this dialogue design an Wizard-of-Oz study was performed. The study had three purposes:

- To evaluate the dialogue design with respect to how verbal and gestural utterances are handled.
- To elicit user behaviour regarding strategies for combining directive and protractive interaction and strategies for managing the movements of the robot.
- To explore the repertoire of gestures and verbal commands with respect to how they are used and what type of variations there are.

The robot used in the study was the Cero robot. In addition to the interface robot on top of the robot cover, a small red lamp and a an ordinary consumer web camera had been placed on the front of the robot. The purpose of this setup was to provide a visual signal to the user that the system had gesture recognition capability (see Figure 4.9).

Figure 4.9 *The environment was laid out in a 2 m X 2 m grid, using white dots painted on the lab floor, 50 cm apart from each-other (for clarity the dots shown in the image have been highlighted manually).*

Dialogue design

The dialogue that was simulated with the prototype should handle different types of input from the user (see Figure 4.10):

- Verbal utterances referring to a small set of directive commands, expressed as restricted language on the same format as on the reference card (see Figure 4.10) given to the users.
- Emblematic gestures corresponding to the directive commands on the reference card for verbal commands.
- A deictic gesture, allowing the user to point to a spot on the floor where the robot should go.
- Verbal utterances of the form "go to ⟨X⟩" allowing the user to specify a specific locations for the robot to go to.

From one perspective the gestures were deliberately restricted to a small set of emblematic gestures. One of the variations that was anticipated was that the performance of the gestures would be different based on how the participant interpreted the depictions on the card.

To give the users the impression that the system had a gesture recognition system we used the lamp that was attached to the robot (see Figure 4.9) as a means of affecting the pace of the interaction. The users were simply told that whenever the lamp was lit the system could "see" the user and receive gesture input.

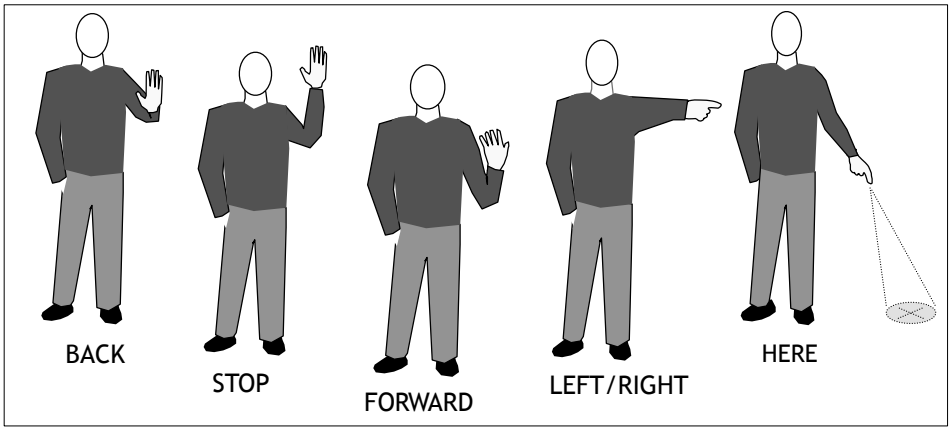
Enaction of the scenario

The role of the wizard was to control the robot so that it moved in two ways:

- Stepwise movement as reactions on the command given by the user (moves are less than 40 cm).
- Driving to a point whenever the user specified a goal-point for the user, for instance, "go to P".

To handle the dialogue the wizard used a dialogue production tool similar to what can be seen in Chapter 3[10]. For each input, corresponding to a gesture or a phrase on the card, there was a specific response, like as in the following example:

[10]Figure 5.6, p. 112



Directive navigation

Halt movement
stop

Navigation word + direction	Navigation word + LOC & GEST
go forward	*go to X*
turn right	*go here & GESTURE*
move backwards	*drive towards Y*
Navigation words: *go, move, drive*	**Location:** <location name>
Direction words: *forward, backward, left, right*	to <location name>
	towards <location name>
	drive to <location name>

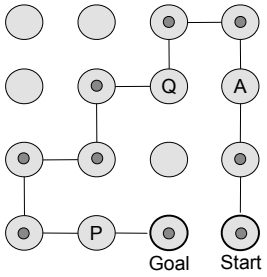


Figure 4.10 *The gestures printed on the card given to the users (top). The card given to the users to be used as a reference during the session (translated from Swedish, bottom left). The task map (bottom right).*

(98) **U:** move forward

(99) **R:** moving forward
⟨ *robot moves forward* ⟩

(101) **U:** ⟨gesture: turn left⟩

(102) **R:** turning left
⟨ *robot turns* ⟩

Participants and scenario instructions

We invited six persons, 2 female and 4 male, to participate in the study. The age of the participants was between 20 and 30 years (the mean age was 25.5 years) and they were all students. They were awarded a cinema ticket for their participation in the study.

The participants were informed about the scenario by the test leader (who was also acting as a wizard and handled the camera). In the scenario they were told they should use the robot to traverse the map (see Figure 4.10, right) and to collect points by passing over white dots that were painted on the floor. This aspect of the scenario was introduced because we wanted the users to focus on solving a task rather than on the interaction situation. We did not keep a score of the points collected by the users, instead every user was told that they had succeeded solving the task.

Regarding interaction they were told the following:

- Moving the robot is performed by using speech and gesture commands.
- The robot understands simple commands which has to do with movement. We did not tell them that there were any difference between the directive commands, like 'turn left", and the goal-oriented commands, like "go to P".
- They were asked to study the reference card carefully as it contained the commands that could be understood by the robot (see Figure 4.10, left).

As this was a Wizard-of-Oz study, the users were not informed that the robot was tele-operated by the test leader. Instead they were told that the test leader would be recording the session using the video camera, supervise the "technical status of the system" and "maintain a log" of the interactions (to explain the sound from clicking the mouse and typing things on the keyboard).

The users were then told to try the robot by commanding it to move to a specific starting point using the phrases and gestures that were presented to them on the reference card (see Figure 4.10). Once they had been given this opportunity the robot was moved to a specific starting position (see Figure 4.10, left) and they were told to start collecting points.

For the user there were then two main ways of solving the task:

- Giving directions to the robot on where to go using the navigation commands: go, move and drive with the parameters left, right, forward and backwards.
- Setting a goal-point for the robot, by telling it to go to location A, P or Q. The users received an instruction which provided them with a clear-cut and simple task: to collect points by moving the robot through a maze painted on the floor.

The maze was constructed so that there would be an obvious way of using "Go to ⟨location⟩" as a short cut directly in the beginning. Since the users did not know what to expect from the navigation behaviour of the robot it was assumed that they at least would think the robot could go in a straight line, but that it would not know how to follow the path (since there were only white dots marked on the floor).

Findings

In the video recordings we observed several interesting patterns relating to how gestures were performed and task strategies. From an evaluation perspective, the overall impression was that all users managed to perform the task using the robot. Most users used the possibility to use the goal-points A, P and Q. One of the users did not use any of these named goal-points.

The starting position of the robot provided an opportunity to collecting two "easy" points by simply saying "go to A" at the start, something which several users took advantage of. It also seemed that some participants used the directive commands to put the robot in a position that would allow the use of the "go to" command with one of the goal points P or Q. It seemed that these users assumed that the robot did not know how the rest of the dots should be passed. A possible strategy in this case would have been to use the three commands to get the robot to the position P from which a simple "go forward" would have put the robot on the

goal point. It seems that the participants assumed that the robot would be able to find a point but that it would go straight to it and not follow any path.

The participants were opportunistic with respect to changing mode of interaction. It seemed that once they had found a working strategy, like using speech only or gesture only, they seemed to stay with that strategy.

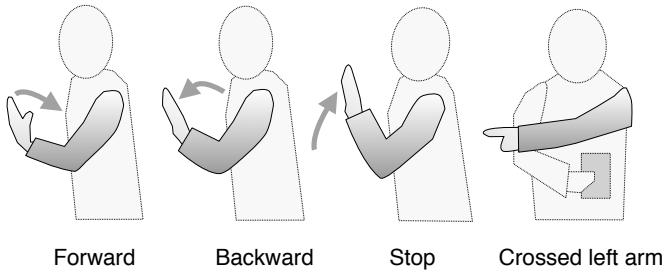


Figure 4.11 *Visualisations of gestures with dynamic patterns shown by the users in the study.*

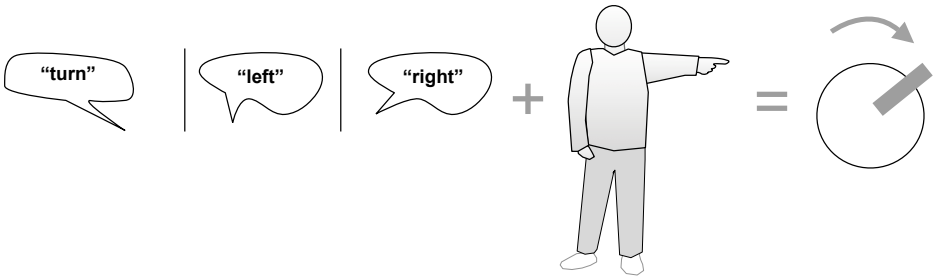


Figure 4.12 *The shadowing effect observed in the study. A verbal utterance that contains an element of turning can be interpreted as a cue for turning and the gesture displays the direction irrespectively of the term chosen.*

Observations regarding manner in which gestures were performed was also an outcome of the study. As users were not equipped with any dynamic description of the gestures that the system was supposed to understand, it was assumed that they would fill in some kind of dynamic movement aspect to the gesture. This could

also be observed in the video recordings. In Figure 4.11 the dynamic parts of the gestures, that are being introduced below are illustrated:

Backward: the static hand gesture (palm up) was sometimes provided with a movement that appeared as if the user was performing a pushing movement, either in a single stroke or as several strokes (similar to a wave).

Forward: some tried to mimic the static hand position whereas others added a waving motion where the fingers and palm was bend towards the body.

Stop: the stop gesture did not give rise to any extra dynamic dimension:

- It seems that the stop gesture is fairly conventional and therefore varies little.
- The phrase "stop" and the stop-gesture was used to provide a way of stopping the robot when a shorter distance was desired.

Left/right: the left or right gestures were performed as shown in the picture with an important exception, sometimes the gesture was performed across the chest (right arms points to the left or vice-versa). It seems that the users expect that the gesture has precedence over the ambiguous symbolic linguistic content, like left vs. right. It seems that the users' display of the "left/right" gestures are used as deictic reference to which way to turn rather than being an instance of a symbolic gesture. It seems that the gesture overshadows the literal meaning of the word "left" or "right". An utterance that has the same meaning is the deictic gesture together with "turn":

- Some users used the left/right gesture in an opportunistic manner. Once the arm was raised the user did not lower it to display a new instance of the gesture. Instead the user merely waited for the signal lamp to light up again in what appeared to be to minimise response time.
- Users got in front of the robot.
- Users seemed to notice the red lamp on the front of the robot since there were very few instances of "repeat" or miscommunication related to non-perception.

4.6 Chapter summary and discussion

In this chapter the design and development process of the robot Cero has been described with a focus on human-robot communication. This chapter has been

focused the process of designing communication for a service robot as well as the design itself. The position of this chapter in this thesis reflects the overall temporal relationship between the Cero project and the Cogniron project. From a conceptual point of view this section can be characterised as an intermediate discussion that will be refined in the following chapters.

Communication design

To provide an answer the first research question posed in this chapter, *"What is an appropriate communication design for an autonomous robot of this type?"*, several topics regarding communication design have been considered.

Challenges of speech recognition

The intention of finding a working solution for speech input to the robot turned out to be a major challenge given the available project resources. In the project we attempted to use commercially available speech recognition for dictation (IBM ViaVoice) and telephony (Nuance) to provide input to the robot. We also tried different microphone setups for far-field speech recognition. The microphones we tested were intended for use in dictation scenarios, not for robotics. Another complicating factor was that we intended to do this in Swedish.

In the wizard studies and the practical testing we observed interesting use aspects of speech recognition technology that were not related to the performance. It seems that on-board speech recognition is accepted and seen as uncontroversial to users. They do not need to see a visible microphone to talk to the robot. When using an offline processing solution for speech input, it appears necessary to give feedback on robot activities, either from the robot itself or from screen-based visualisation. On the one hand wearing a head-set microphone may be experienced as intrusive by users. On the other hand, speech not directed to the robot such as talking to a bystander or answering the telephone is picked up by microphones, something which also has to be considered in the design.

In retrospect we underestimated the complexity of the problem of making on-board speech recognition work. Given the state of the art in speech processing and robotics, even today it is not clear that more economical resources would solve the problem.

Dialogue design

The dialogue model developed in the Cero project focused on handling specification of task-goals. These goals could either be characterised as protractive or directive. Protractive goals concern services that are extended over large distances and lasts long time. Within in the context of the project protractive goals concerned going to locations, delivering and fetching objects from various locations. Directive goals concern actions that the robot can carry out directly such as small movements or information-oriented actions. In the context of the project, directive goals concerned going forward, turning left or right and stopping.

To handle protractive goals the dialogue strategy used was cautious. This meant that the system attempted to ground actions before embarking in missions. Directive goal specifications were assumed to be grounded once they had been received and carried out immediately. When negotiating and managing long-range (protractive) goals using the dialogue system the responsibility for plans created lies with the system. Plans specified using the system needs to be:

- possible and fully specified before being committed to by the system, and,
- they should lead to a service being carried out.

Task-oriented actions for short term or information goals can be executed reactively without using a cautious grounding strategy. When actions that are specified means that the system need to override it's safety systems, like the obstacle avoidance, the plans carried out with the system are no longer represented by the system. This means that there is no way of detecting fatal actions. In practise the robot is being operated directly by the user as part of her plan. Since possibly dangerous actions can be carried out with the system, allowing an optimistic grounding strategy in the system requires careful consideration. By including directive commands in a system that comprises a non-robust speech interface may lead to undesired consequences. It is possible that the system picks up sounds from the environment that translates into directive commands. *In a real life scenario users needs to be informed about this.*

To provide a natural-like dialogue, the dialogue model incorporated both task-oriented dialogue (for protractive and directive goal specification) and feedback. Feedback needs to be given rapidly and at relevant points during dialogue. In the Cero system the feedback model by Brennan & Hulteen (1995) was used as an

inspiration for the feedback model of the dialogue system. Feedback can be given in several modalities in parallel, through gestures from the Cero interface robot, by verbal utterances, audio signals and by the robot movements.

Evaluation approach

To answer the second research question in this chapter, "How can we approach evaluation of the quality of interaction?", different types of evaluation approaches have been used.

The Cero project can be viewed as a research oriented learning process rather than a user-centered product development. The technical assessment phase of the project meant that we performed several practical tests with implemented system components, both specific component and integrated prototypes. We also carried out Wizard-of-Oz simulations to gain first hand experience and to collect data that would allow post-hoc analysis of video recorded trial sessions. One thing that should be stressed is that engaging a skilled industrial designer early in the design process had positive benefits. It appears as the form factor of the robot is important to provide a situation which users experience as realistic.

The Wizard-of-Oz simulation technique is a method that is best suited to the initial and explorative phase in the design process. The technique can give insights early in the design process, but cannot be the first step since the technique has a mandatory requirement that a prototype design has to be conceptualised.

In order for wizard operators to successfully enact the behavioural characteristics of an interactive robot, the wizards need a combination of model capabilities that form the services provided by the system and an algorithm which provides the constraints to determine with what level of competence the system can be portrayed to potential users.

The immediate outcome of the Wizard-of-Oz studies performed in this project was the opportunity for designers to first hand experience of acting as the system. By collecting data we could also put users in a situation where they experience situation that represented a scenario of future use.

The initial Wizard-of-Oz study pointed out the need to provide sufficient constraints both for the users and for the wizard operators. Even if the system that was enacted was unrealistically competent, given the state of the art in human-robot interaction, we observed phenomena that informed the future design. We saw it

necessary to provide a design that afforded a heading for the robot. To increase this impression we also introduced the notion of an interface robot.

The physical character of human-robot interaction, experienced in the Wizard-of-Oz study, led to decisions concerning the dialogue design, namely that grounding strategies were needed to handle goal specification.

The second Wizard-of-Oz study was focused on strategies for handling the robot in a small space, using specification of directive and protractive goals. Regarding user strategies, it seemed that when using a named location a protractive goal was preferred over specifying several directive goals.

The interaction in the second Wizard-of-Oz study was constrained by giving the users a reference card with the verbal and gesture commands that the system could handle. It seems that when given static gesture descriptions, users fill in their own dynamics. Another phenomena that could be observed was that it appears as users assume gestures to have precedence over spoken utterances.

The practical tests with the dialogue system and the primary user, and the video-recorded session with the expert user illustrates the difficulties of creating a usable system. The poor performance of the speech recogniser stressed the need to focus research on miscommunication. In both sessions the lack of feedback or ill-timed feedback led to miscommunication. In the practical test with the primary user, there was no clear model for how the interaction should be carried out, something which led to meta-discussions that was picked up by the system. As there was no robot available, it was also unclear to the user what services the robot could perform and consequently what should be said.

In the session with the expert user, both the services available and the dialogue model were known to the user. Instead the miscommunication had to do with speech input and timing of feedback. Another thing which was observed in the session with the robot and the expert user was that small movements of the robot were interpreted as evidence of understanding. This stresses the importance of testing with a robot physically present.

Focus shifts in the design process

By developing the Cero we were able to gain invaluable first-hand experience of human-robot communication in a realistic setting. The final result, the Cero robot, turned out to be something quite different than what was expected in the initial

phase of the project. Because the development process was research-oriented rather than product oriented, we could diverge from the initial goals concerning human-robot communication. From a communication point of view there were several focus shifts which can be seen a sign of a more informed view on human-robot communication:

- Initially we addressed problems related to natural language understanding. During the project focus shifted to investigate challenges of providing timely and relevant feedback based on robot perception.
- At the project start human-robot communication was viewed as a primary verbal activity, but we realised that multimodality, including verbal and gestural utterances, spatial orientation and physical movements need to be addressed in order to understand human-robot communication.
- The interviews, questionnaires and initial conceptualisation of the system provided motivation for embarking on a practical design activity. Once prototypes were available, the results of the initial interviews and questionnaires had a very limited value. This can be characterised as a shift from conceptualisation of hypothetical scenarios to reflective practise of prototyping.

In the next part, describing the subsequent work performed in the Cogniron project, the challenges provided by these focus shifts will be explored further. By analysing data collected in user studies with simulated systems I have studied miscommunication, spatial aspects of interaction and how feedback can be supported based on the robot's perceptual capabilities.

CHAPTER 5

Developing a Corpus for Human-Robot Communication

The purpose of this chapter is to describe the development of a corpus which was used to evaluate and analyse human-robot communication in relation to the Cogniron project. This chapter describes the Wizard-of-Oz study conducted in the initial phase of the project. The study was performed with two goals in mind:

- The first goal was to evaluate a specific dialogue design as a step in the project's iterative development process. Although the particular result of this evaluation is interesting, the main benefit was to inform the design and was communicated directly within the project.
- The second, and for this thesis most relevant, goal was to collect data to develop a corpus to allow in-depth studies of various topics[1] relevant to the project and human-robot interaction in general.

In the chapters following this one I will focus on three aspects of human-robot communication which are all based on analyses of corpus data: miscommunication (Chapter 6), the role of perception and contact feedback (Chapter 7) and spatial influence (Chapter 8). Figure 5.1 illustrates how the evaluation activities and focused analyses described in this thesis are related to the research process in the Cogniron project.

[1] Work by Hüttenrauch et al focusing on Task Strategies (Hüttenrauch et al., 2006b) and Spatial Relationships (Hüttenrauch et al., 2006a) have also been based on the data collected in this study.

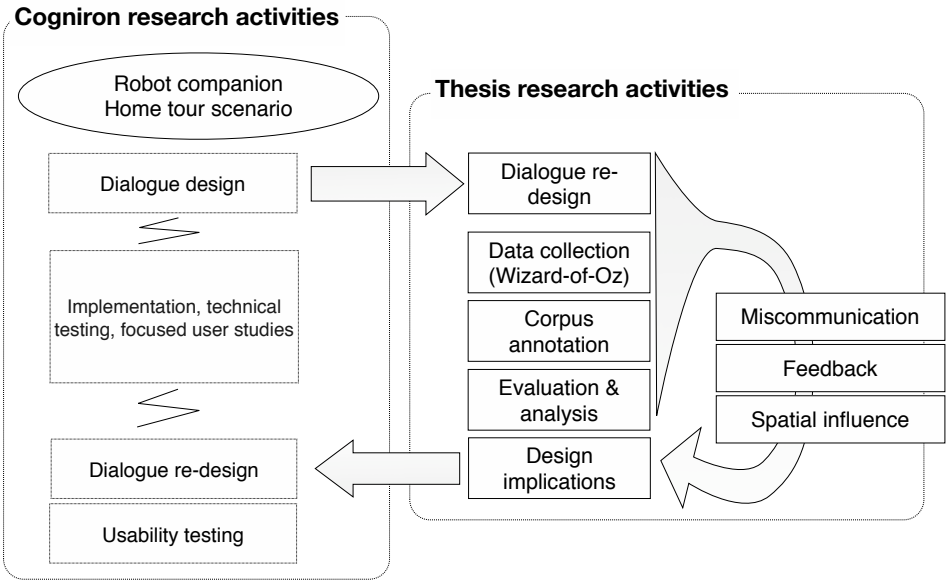


Figure 5.1 *The thesis relation to the work performed in Cogniron*

5.1 Previous approaches to corpus data collection

Several initiatives to collect corpora for multimodal interfaces have been reported in literature (Knudsen et al., 2001; Schiel et al., 2002), but few are targeted for robotics (Bugmann et al., 2004, 2001; Wolf and Bugmann, 2005). Koide et al (2004) have collected and analysed interaction statistics to investigate human reactions to specific robot behaviours. Other uses of corpus data include observations of user behaviour, for instance, gaze behaviour, to evaluate human engagement in interaction (Sidner et al., 2004).

Perzanowski et al (2003) collected corpus data to investigate spatial language in a scenario for a handheld computer and speech input. Lauria et al (2002) collected route descriptions for robot navigation similar to the Map Task corpus (Carletta et al., 1997). The practical setup in the collection was not a proper wizard setup instead participants were told they were interacting with a human operator hidden in another room. The small robot used in the setup was located in a physical model world, with the user standing beside it. The participants were instructed to address the operator, sitting in another room, as if the operator could see the video image

from a camera mounted on the robot. Using the corpus a grammar could be constructed, specifying the primitive procedures used by the participants to describe the route for the robot (Lauria et al., 2002).

5.2 The Cogniron Home tour scenario

One of the central themes of the Cogniron project was to investigate a scenario where a cognitive robot is introduced to its operating environment by a human teacher. This scenario was called the *Home Tour*. In this scenario, the robot discovers and builds up an understanding its environment. The human teaches features of the environment to the robot in a continuous communicative process:

> "A human shows and names specific locations, objects and artifacts, to the robot. The robot can engage in a dialogue in case of missing or ambiguous information" (Cogniron, 2003).

At an abstract level the user and the robot are engaged in what can be characterised as a co-operative service discovery and configuration. In other words, the user and the robot are engaging in a joint effort to share relevant knowledge about the environment. This means that the user is able to discover what the robot can do and to configure it by actively providing information about:

(i) the *artifacts* present in the environment (like objects and locations) and,

(ii) the *actions* which the robot can perform related to these artifacts.

5.3 Wizard-of-Oz study III: Data collection for evaluation of the Home tour

We carried out an activity analysis of the Home Tour scenario and used this as a starting point when we planned the data collection study. In this section I will introduce this analysis and how it was used when we re-designed the dialogue and subsequently the study used to collect data.

The role of the user

The role of the user in the Home Tour scenario was to act as a guide the robot in an environment containing objects and locations that the robot could recognise. The user's main task then is to introduce herself to the robot and to show it objects and locations. To give the user a sense of closure we have also added a validation task.

This means giving the user a possibility to try the functions the robot is supposed to have learned. It is also possible to end the interaction with the robot by using the conventional means of closing an interaction, for instance by saying "Good bye".

In the scenario there five types of activities were supported by the robot:

- *Introduction*. The user may introduce herself to the robot. Directly after the introduction the robot will state that the user has been recognised and remembered.
- *Demonstration of objects and locations*. The user shows and names objects and places to the robot using speech and deictic gestures.
- *Activation of the following behaviour*. Using the following behaviour is the intended way of controlling the movements of the robot. The follow behaviour of the robot is used to position the robot in the experiment area to allow the user to present an object or location.
- *Validation of the learning process*. The user is able to find out if the robot has learned objects and locations by requesting the robot to go to locations and to find objects in the environment.
- *Closing the interaction*. The user may close the interaction by using a spoken command, for instance a parting phrase like "good bye".

The activity descriptions work both as an aid for the wizard and a constraining factor for the scenario, meaning that we bring realism into the situation of use. Realism is something which is important for the overall quality of the collected data (Maulsby et al., 1993).

Re-design of the Biron dialogue

In order to be able to simulate a working prototype we used the dialogue patterns (see Figure 5.2) provided by one of the partners in the Cogniron project. Using these dialogue patterns as a starting point we re-designed the dialogue with the twofold goal of creating a simulated prototype that first of all would be perceived as complete, or habitable, by the user, meaning that where there are no, or few, points where breakdowns occur because there is no "model" for handling a specific phenomena. Secondly the goals of the re-design was to familiarise the wizards with the dialogue model and the actions the system was supposed to accommodate.

The first goal is not intended to mean that the user should be able to utter an arbitrary command to the system and expect the robot to understand it, but rather that there is a relevant help message for those cases when there is no obvious response for the particular error event, according to the task model, making it possible to recover from the situation at hand. This means that sometimes the wizard must make the robot appear almost uncooperative, to stay within the boundaries of the simulated dialogue. How this was handled can be seen in the excerpt from the dialogues that were collected in this study:

(104) **U:** good work robot
(105) **R:** What did you say
(106) **U:** good work
(107) **R:** I did not understand please rephrase
(108) **U:** you have accomplished the task
(109) **R:** What did you say
(110) **U:** follow me

Example 5.3.1: An excerpt from the Corpus collected in the Wizard-of-Oz study. The underlined utterances show responses from the robot intended to constrain the behaviour of the user.

The process of re-designing the dialogue started by investigating the explicit dialogue patterns, supplied by our colleagues in the project, for the functions the system would handle: *Greetings* and *Closing*, *Person following* and *Object and gesture detection*.

The manner in which we have reworked the dialogue patterns results in a task-oriented dialogue model together with one or more patterns for error handling. The task-oriented model handles the behaviour the system displays under normal (positive) circumstances. States which can be associated with some error state and should be detected by the wizard operators) with the error handling dialogue for a particular task. One such example is depicted in Figures 5.2 and 5.3, where the original dialogue design considered for implementation in the BIRON[2] robot,

[2]This robot was one of the robots used for the Cogniron Key Experiment (KE1) and is described in (Haasch et al., 2004)

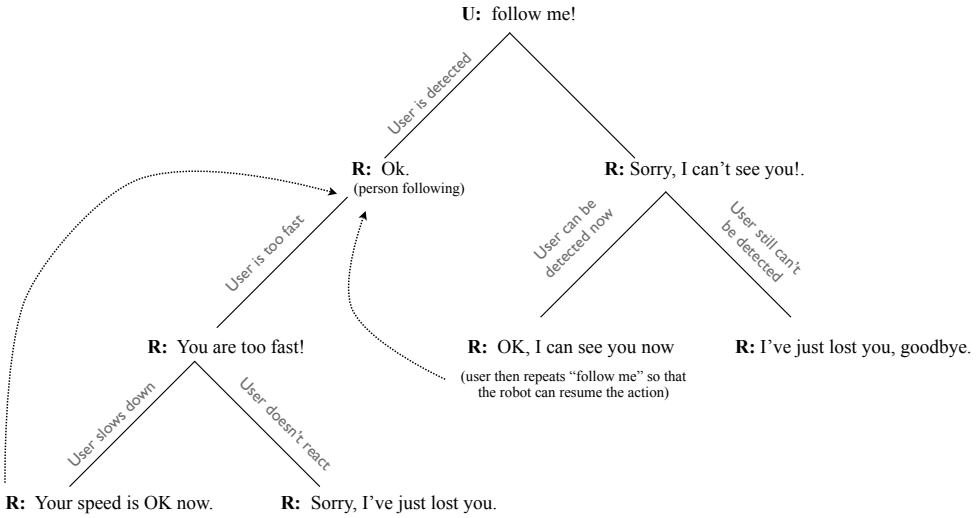


Figure 5.2 *An example of a state-based-dialogue model used as a background model in the redesign process leading towards the dialogue tested in the Wizard-of-Oz study. This dialogue shows the state-chart for the dialogue intended to activate the Following behaviour of BIRON, cf. (Li, 2007), for an in-depth description of the dialogue design implemented in the BIRON dialogue.*

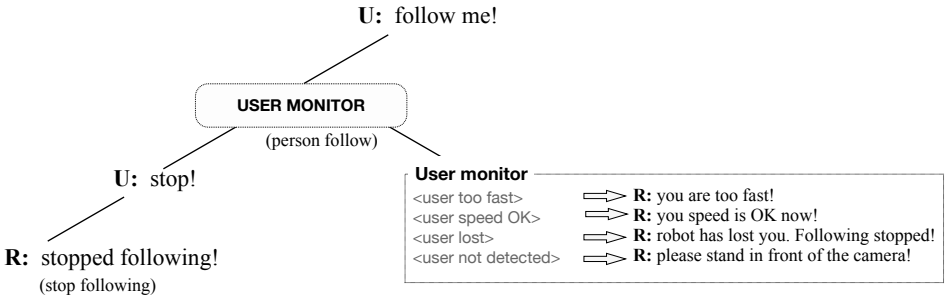


Figure 5.3 *The re-design of the dialogue shown in Figure 5.2. The system states identified in the original state-chart have been replaced with a set of situations which the wizard-operator can detect.*

shown in Figure 5.2 is re-designed to better suit behaviour the wizard can simulate, or enact, through the robot platform used[3] on the study.

[3] In the experiments we used a robot which was similar to the robots used in Cogniron, an ActivMedia Peoplebot (see Figure 5.5).

Robot behaviour

We also considered the robot behaviour in the design. The system that we were investigating in the course of the Cogniron project was intended to have a Person Attention System (Haasch et al., 2004). The version that was used at the time could only detect audio input from a user that had been detected using a face recognition algorithm. This constraint was due to limitations in the automatic speech recognition system. We believed that the participants would not accept a robot that worked in this highly restrictive manner. We also judged that it would be hard for the wizards to maintain this constraint. Consequently we allowed the system to perceive sound independent of the position of the user.

Another design decision that was taken during the setup of the trial was to bring camera movements into the robot design. Initially we explored the possibilities to capture camera images. At this point we viewed the camera as a data source among others, but we quite soon realised that the movements of the camera also would affect the way the robot was perceived by the participant. We decided to assume a model that incorporated a moving camera (for capturing) and camera gestures with a communicative intent. When a spoken prompt was issued, it was accompanied by a camera gesture displaying a gesture that makes the robot "look up" slightly towards the participant. This is an example how the findings in the Cero project (Chapter 4) had an impact on later research.

In general the follow behaviour worked as in the description. The movement pattern of the robot behaved like a rubber band due to the reaction time of the wizard and the attempt of assuming the minimum distance of approximately one meter. Mostly the navigator wizard tried to follow the participants by turning directly towards them and then move in the same direction as they were walking. At some points the follow-wizard departed from this main pattern and instead used a deliberative model of placing the robot:

(i) Assuming a position that would have a desired effect on positioning of the user, for instance, when the robot is placed so that the user will not stand in a position that will make it possible to reveal what is going on in the wizard booth.

(ii) Positioning that will facilitate object recognition based on the condition that the wizard can anticipate what object is about to be defined.

The initial dialogue pattern for showing objects assumed a model where the robot had a touch screen. This dialogue was re-designed to handle a camera based object recognition model. The original pattern contains different prompts that were specific to the touch screen (like *"please click on the ⟨OBJECT⟩ on my touch screen"*). We used a small set of prompt patterns in the sessions, assuming that the object recognition system would "work" unless there was some obvious error state, like several objects visible or no object visible in the camera view. The prompts[4] that were used for showing an object during the sessions, are listed below:

(111) **R:** Found one ⟨OBJECT⟩

(112) **R:** I do not know that object

(113) **R:** Found several objects

(114) **R:** Rearrange the objects please!

Example 5.3.2: Prompts for handling the show objects task

Roles and Work task for the Wizard operators

One assumption for the whole task of controlling the robot's movements, its camera and its speech capability was that this was a job for more than one wizard. After some consideration we decided that the one wizard should control the movements of the robot platform and the other should control the dialogue behaviour. After a technical assessment of the platform capabilities we also added control of the on-board camera to the wizard tasks. The division of the wizard role is not only made on the basis of technical considerations, but also reflects the conceptual difference between moving and communicating. Hence one wizard role is that of the "Navigator" and the other is the "Communicator". In the setup the Navigator wizard also acts as test leader towards the participant. The Communicator wizard acted as "Technician" towards the participant. The test leader and technician roles could be switched. An overview of the wizard roles is shown in Figure 5.4.

Defining the wizard task as the two subtasks: navigation and communication has been the preferred model of others that have attempted at simulating multimodal human-robot interaction. For instance, Perzanowski et al (2003) used a

[4]The ⟨OBJECT⟩ placeholder was expanded to a set of phrases containing the known objects in the dialogue tool

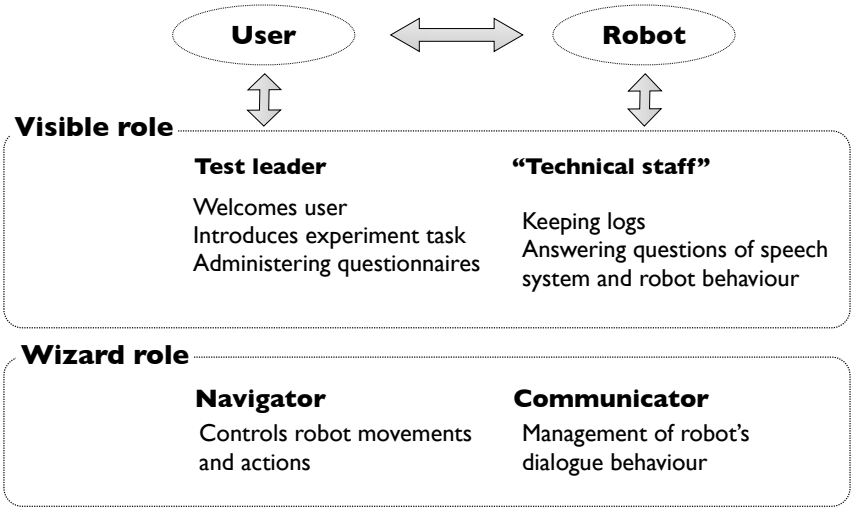


Figure 5.4 *The wizards' visible and covert roles*

this division of labour in a pilot study to collect multi-modal data. One wizard was controlling the navigation of the robot; the other acted as the robot's speech interface using a headset microphone attached to a sound modulator to produce a robotic voice. Two wizards were also employed in the study described in Chapter 4. During the sessions one wizard was responsible for the physical movement of the robot and the other provided dialogue capabilities by playing prompts using the on-board speech synthesiser (Perzanowski et al., 2003).

Experiment setup

The wizard scenario was set up in an experiment environment called the "Living room" located in the robot lab at KTH[5](see Figure 5.5, right). This is an office room (5 X 5 metres) furnished with typical Swedish furniture[6] to resemble a living room in a Western European home. Other than furniture, the room was furnished with a set of everyday objects. On a table in one corner of the room a screen was put up to cover the two computers that were used by the wizards. We wanted to prevent the participants from directly seeing what the wizards were doing.

[5]The Computer Vision and Active Perception Laboratory (CVAP)
[6]Mainly from IKEA

Figure 5.5 *The robot used in the Wizard-of-Oz data collection (left). The room where the Home-tour scenario was enacted (right).*

The robot

The robot we used for the trials was an ActivMedia Peoplebot[7]. The robot (see Figure 5.5) was equipped with four visible microphones, two of which were attached to metal wires on the top of robot to avoid picking up motor noise from the robot platform. Two other microphones were less prominent but yet visible to the participants. The robot also had a pan-tilt video camera providing a simple gaze mechanism for the robot. The gripper was concealed by a digital sound recorder used for the collection of on-board sound. One prominent feature of the lower part of the robot was a laser range finder with the (standard) clear blue colour. On the upper and lower part of the robot two sets of ultrasonic sensors were attached. The ultrasonic sensors were switched off during the experiment to reduce the noise from the robot. Still the level of noise from the fan of the robot's on-board computer and the motors was considerable.

Participants and procedure

Initially we performed a formative pilot study with a few staff members in order to fine tune the setup. In the next phase we recruited 22 participants among students

[7] http://www.mobilerobots.com/

on the KTH campus. This means that there is a bias towards well-educated and young people in the study, but since the goal of the study was primarily explorative we accepted this circumstance. Upon arrival the participant was greeted by the test leader and offered a cup of coffee. Then the test leader informed the participant of the purpose of the study, without revealing that the wizards were controlling the system. Instead the wizards were described as "technicians" with the purpose of controlling the technical setup and making "on-line annotations". During the trial there were three researchers present; one acting as test leader/navigator; one acting as communicator; and one acting as observer. During the setup the observer was positioned in one of the sofas taking notes. After the introduction the participant signed an agreement giving consent to storing of personal information and then was instructed to read the written instruction about scenario and the task.

After the participant had finished reading the instruction the test leader addressed any questions or requests. Then the test leader gave the following demonstration standing in front of the robot:

(115) **TL:** Hello robot!
(116) **R:** Hello I am ready
(117) **TL:** Follow me
⟨ *robot follows TL* ⟩
⟨ *TL stands in front of book shelf* ⟩
⟨ *TL points at book* ⟩
(121) **TL:** This is a book
(122) **R:** Found one book
(123) **TL:** Go to the battery re-charge station

Example 5.3.3: The demonstration given to the participants before starting the test session.

The dialogue production tool

To provide multi-modal dialogue capabilities for the robot the Communicator wizard has a tool (Figure 5.6) that provides output from a large set of phrases. Since the dialogue interface simulated here is intended to handle task-oriented dialogue we have assumed that phrases may have two functional types: task-related and general

feedback. This is reflected in the interface where the left table holds task-related phrases and the right table holds feedback phrases.

To handle phrases containing locations or objects we have added columns to hold objects and locations. When a task-oriented phrase containing a type marker (for instance, *"LOCATION"*) was selected, a dialogue window containing the set of expanded phrases for the possible location is displayed (see Figure 5.7). This makes it possible to produce hundreds of phrases with a few mouse clicks.

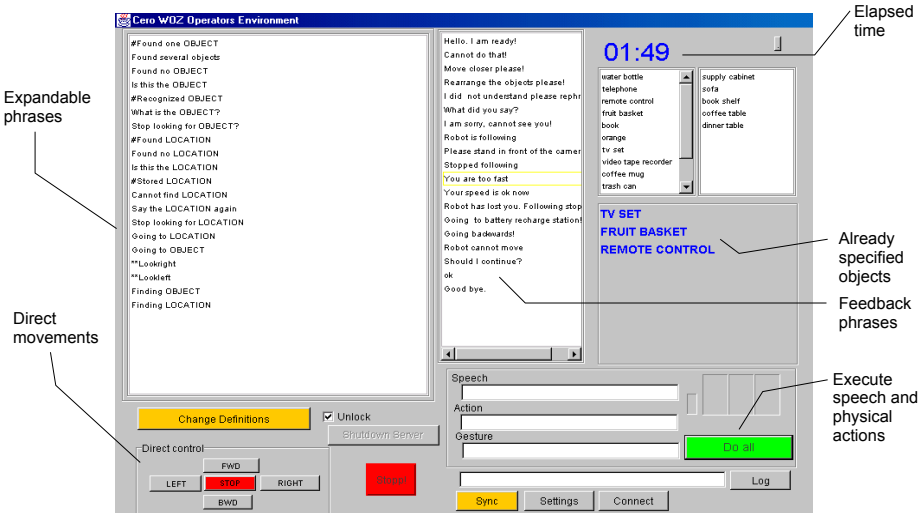


Figure 5.6 *The Dialogue production tool. The image shows the different functional elements used to control the dialogue.*

The wizards also have access to a list showing objects that have been mentioned during the session. The list was added to the interface after the pilot sessions. A stop watch timer was also added to keep track of the length of the sessions. The tool also contains fields for commands to produce simultaneous robot actions, for instance by sending a move command while letting the robot say "moving forward". This feature was used to provide some camera movements corresponding to communicative feedback (for instance by looking slightly upwards when saying "Hello").

Using the navigator tool the wizard is able to directly control the robot's movements using a standard type gaming joystick. For this task the most important

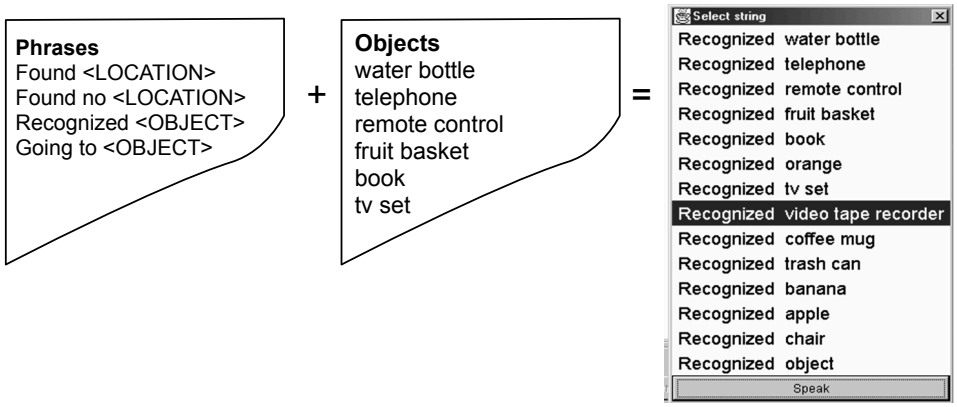


Figure 5.7 *Phrases containing objects and locations can be expanded to rapidly produce complete sentences without the need to type the rest of the phrase.*

feedback to the wizard is provided by directly monitoring the robot itself. The on-board camera image also gives some information that can be used to decide where the robot is looking.

Data Collection

In order to be able to analyse the study from different perspectives data of various types were collected. First of all video from the overall scene was recorded, using a MiniDV camera. This camera was placed on a tripod behind the wizards' screen and handled by the Communicator wizard. This camera recorded video (MiniDV) and audio and was operated by the dialogue wizard. The camera was equipped with a wide angle lens in order to capture as much of the scene as possible and minimising the need to pan the camera during the trial sessions. We also collected images from four network web-cams, placed in each corner of the room. The purpose of collecting this data was to get an overview of the scenario, and to figure out what went on when the main camera was occluded. The frame rate of the network cameras was approximately one frame per second, depending on network traffic and load of the servers where the images were stored. Audio from two different sources was collected: the sound from the wizard's video camera and the sound from the stereo microphones placed on top of the robot. The on-board sound was recorded on a digital audio recorder placed on the robot gripper. The microphones were facing forwards, and sat on top of the robot and with a distance

of 40 cm between them (see Figure 5.5). The microphones were also arranged to minimise the sound from the robot platform.

Each interactive session lasted about fifteen minutes and contained between 60-100 verbal and gestural exchanges between the participants and the robot. The sessions are annotated on the utterance level, defining an utterance as something which to the transcriber seems to be a coherent sequence of speech. This means that sometimes the word "robot" followed by a significant pause and a command is treated as two separate utterances (for instance, "robot" and "follow me") . In other cases the similar constructions, starting with "robot" is taken to be one utterance, like "robot follow me". An automatic log was kept whenever the wizard made the robot speak. By inspecting a spectrogram view of the on-board audio file the offset between the time of occurrence for the first robot utterance in the recorded session and the log time of the corresponding command in the log was established.

In the next session I will discuss how communicative acts were modelled in the corpus data.

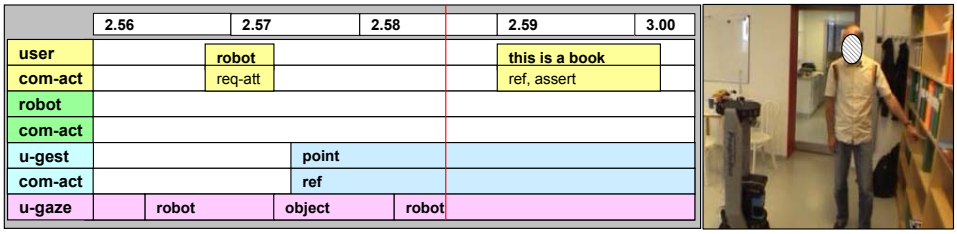


Figure 5.8 *Different corpus data visualised as a (slightly simplified) score annotation similar to the visualisation in the Anvil tool (Kipp, 2004) that was used for the annotations.*

5.4 Annotation of multimodal communicative acts

For the corpus an annotation schema was developed that was constructed to represent human-robot communication. Both the diversity and the large quantity of data provide challenges when it comes to annotate data in a way that is useful for analysis. The data need to be annotated in a way that enables search and visualisation of events, based on synchronised annotations over several data modalities.

Forward-looking functions		
Action-Directive	Info-request	Assert
Reference	Reassert	Offer
Commit	Request repair*	Opening
Closing	Request-Contact*	Attempt-Contact*
		Other forward-looking function
Backward-looking functions		
Report-Action*	Report-Action-Fail*	Accept
Signal non perception	Signal non understanding	Acknowledge
Reject	Provide-Attention*	Provide-Contact*
		Other backward-looking function

Table 5.1 *Forward-looking and Backward-looking functions. The functions marked with '*' are extensions to the DAMSL coding schema.*

One aspect of human face-to-face communication that need to be considered when modelling communication, is that utterances may have multiple communicative functions. Conversations are also carried out simultaneously in several communicative tracks. In the model of grounding proposed by Clark (1996) primary tracks are used for basic communicative acts, like assertions: "The weather is nice", and secondary tracks are used for meta-communicative acts (from a task-perspective), like feedback of various kinds, for instance related to perception: "I hear you".

When shifting from a domain where the conversation concerns information management, the topic of concern is shared as a virtual object rather than physical. In the Map Task (Carletta et al., 1997) or the TRAINS corpus (Allen and Core, 1997), the grounding process concerns the mutual understanding of a map given to the participant or the joint creation of a plan for transporting goods. In the robotics domain the topic of conversation is the actions of the robot itself. To account for contributions related to robot actions tag-sets like (Allen and Core, 1997; Carletta et al., 1997) need to be extended.

Annotation of communicative functions

Apart from the annotations related to spatial orientation and interaction episodes (Hüttenrauch et al., 2006a,b) we developed a coding taxonomy to represent communicative acts. The primary purpose was to label communicative acts expressed using verbal and gestural utterances. The goals of this annotation was that it should be:

- Neutral with respect to specific theories of interaction, meaning that it should support different analyses. Either by allowing for transformation to other categories or by being possible to extend.
- Incrementally usable meaning that it should be possible annotate data and to carry out evaluations step-wise and in parallel. The goal is to allow for intermediate results to inform dialogue design carried out within the project, while at the same time allow for analyses with respect to research on human-robot communication.

The categorisation described in this thesis has been included in a slightly modified version in the broader taxonomy developed within the Cogniron project (Otero et al., 2007).

The coding taxonomy can be viewed as a multimodal extension of the DAMSL[8] coding schema (Allen and Core, 1997). The labels used to describe communicative functions that to a large extent are defined in (Allen and Core, 1997). The DAMSL framework has been extended with categories related to physical action (Request-Repair, Report-Action, Report-Action-Fail). In Table 5.1 the list of communicative functions that has been used to annotate data in the corpus used for the analyses in the following chapters. For each utterance a set of forward- or backward-looking functions are identified. The effect of performing a communicative act with forward-looking function is prospective and concerns the context following the tagged act (like prompting for action). Backward-looking functions are related to previous context or acts (like feedback). It is also possible to provide a comment to give a description of the behaviour and try to consider possible interpretations regarding the informational value of the interaction.

[8] Dialog Act Markup in Several Layers

In addition to the schema above contributions related to the management of attention and willingness to interact with the categories Request- and Provide-Attention, and Request- and Provide Contact. Management of contact and attention can be performed using different modalities (Allwood et al., 1991). This draws on findings by (Allwood et al., 1991) and extends the schema adopted by (Gill et al., 2000) who annotated the body move category Attempt-Contact.

While verbal utterances are transcribed using conventional orthographic form. Gestures are transcribed only with respect to their communicative function. This means that the set of extended DAMSL annotation categories, displayed in Table 5.1, is used to describe communicative functions of both verbal and gestural utterances.

If there is a related non-verbal gesture we annotate this relation and its type. If there are more non-verbal gestures relating to this utterance, we annotate the most prominent relation according to the following schema modelled after Kendon (1997):

Co-produced (special case of speech and gesture) – the gesture and the verbal utterance are together forming a meaning. This is further specified by a the following aspects:

Content-aspect – we annotate if the gesture provides content emphasising or influencing the meaning of the utterance. The content itself is not coded.

Deixis – we annotate if the gesture is co-produced and provides a deictic reference to a domain object. The object referred to is not coded.

Conjunct – the gesture is produced alongside the speech without providing lexical meaning (for instance, gesticulation alongside intonational patterns). This aspect is *not* coded.

Gesture reference – the reference between a verbal production and a gesture annotation is annotated (by linking two tracks).

Communicative function – the communicative function of an utterance is expressed as a set of functions taken from the range of labels available in the tag-set, listed in Table 5.1.

During the analysis of spatiality we also wrote down observations on events in the session. These text descriptions have been time aligned so they can be used as links to specific points of interest in the data. Answers to questionnaires administered to participants concerning their attitudes towards the system are also available.

SESSIONS AND PARTICIPANTS		
Participants Data from 22 participants were collected. The participants were mainly students (9 female, 13 male, ∼24 years old).	**Sessions** The duration of the sessions were about 15 minutes. No session was shorter than 10 minutes and none longer than 20 minutes	
COMMUNICATION ANNOTATIONS		
Verbal utterances Transcribed with regular orthography. Communicative functions according to the extension of DAMSL	**Gestures** The gestures were transcribed with respect to their communicative function according to the annotation schema.	
TASK ANNOTATIONS		
Task events Shifts between tasks: follow, show, locate, etc. Naming objects and locations.	**Task-annotations** Text descriptions of notable actions and events and behaviour, non-time aligned, related to utterances. (cf. Hüttenrauch et al. 2006b)	**Positioning data** Hall distances and F-formations, data from laser range finder (cf. Hüttenrauch et al. 2006a)
BACKGROUND DATA		
Task-descriptions Descriptions of user and robot behaviour	**Questionnaires** Questions and answers regarding attitudes and background data of participants	**Post-session interviews** Questions and answers regarding general impressions of communication and interaction with the system
MEDIA		
Video MiniDV converted to .avi files	**Audio** On-board (stereo, 16KHz), Camera sound MiniDV	**Still images** Images collected from cameras in each corner of the room (∼1 fps). Images from on-board camera (∼1 fps)

Table 5.2 *An overview of the data collected in the first phase of the Cogniron project.*

5.5 Chapter summary

This chapter has introduced a corpus that was collected to study human-robot communication. An overview of the corpus data (focusing on the data relevant for this thesis) can be seen in Table 5.2. To collect data we performed a Wizard-of-Oz study with two goals: evaluation of a specific dialogue design and data collection to allow for in-depth studies of human-robot communication. The specifics of the evaluation of the dialogue was not treated within the scope of this thesis. These challenges related to data collection were introduced and discussed:

- *Re-design of the dialogue based on state-charts available in the project.* The design was redesigned with the goal to create a dialogue that was believable with respect to interaction and that could be managed by a human wizard operator.
- *Enactment of the system capabilities.* To enact the system capabilities the task was split into the two operator roles "Navigator" and "Communicator". To the participants these persons were introduced as "Test leader" and "Technical staff" something which made it possible to enact the scenario in a single room.
- *Data annotation.* The data from the study were annotated using a multimodal extension of the DAMSL annotation scheme. This extension incorporated communicative actions related to contact and perception as well as gesture relations modelled after Kendon (1997). Gestures and verbal utterances are analysed as multifunctional communicative acts, meaning that the communicative function of an utterance, not the means of production, guides the assignment of the category.

The corpus data collected in the study will be used in the in-depth analyses of human-robot communication described in the following chapters, analysing miscommunication (Chapter 6), perception and contact feedback (Chapter 7) and spatial influence (Chapter 8).

CHAPTER 6

Miscommunication Analysis in the Design Process

This chapter describes how analysis of miscommunication can be integrated in the process of designing communication for a service robot. The analysis was carried out based on the corpus on human-robot communication described in the previous chapter. The analysis of miscommunication described in this chapter is intended to answer the research questions regarding the type and characteristics of miscommunication and ways of preventing or reducing it in human-robot communication. The way this was approached was to analyse the corpus data to identify trouble-spots – sequences of interaction that displayed signs of miscommunication. In the following we will introduce the types of miscommunication occurring in the sessions we analysed, and discuss design implications focused on how miscommunication can be reduced or prevented. First I will introduce some relevant notions on communicative quality.

6.1 Communicative quality

In this thesis communicative quality is important in two different ways. First of all normative approaches to communicative quality, like frameworks based on gricean maxims, design guidelines, and practical conventions, provide us with the means of improving communicative quality, for instance by preventing errors and the possibility of ending up in what can be characterised as a severe state of misalignment during interaction. Second, models and theories that explain causes and symptoms

of miscommunication are important when communicative quality in a system that is being is evaluated.

One of the most influential theories used to explain and assess the mechanisms of communicative quality is perhaps the theory of Grice, who proposed the cooperative principle as a social protocol, shared between participants of a conversation (Grice, 1975). By assuming the cooperative principle and the four maxims (see Section 2.2, p. 19) we can reason about pragmatic phenomena in conversation in terms of communicative quality.

Miscommunication can be defined as a state of misalignment between the mental states of agents involved in communication (Traum and Dillenbourg, 1996). Either the speaker fails to produce the effect intended with the communicative acts issued or the hearer fails to perceive what the speaker intended to communicate. Analysis of miscommunication is sometimes referred to as "breakdown analysis", but a breakdown is an almost hypothetical extreme case in a wide spectrum of possible miscommunication. Instead of using the term *'breakdown'* we refer to this as a severe state of misalignment between participants in dialogue. Multimodal interfaces are far from perfect. Many challenges can be contributed to the lack in robustness of components for speech and gesture recognition. This challenge should be approached along two, equally important, dimensions: by improving the performance of recognition components through technological development, and by providing dialogue models that handle conversational errors and "gracefully" reduce miscommunication (Bohus and Rudnicky, 2005). The work presented in the following is limited to analysis and re-design[1] of dialogue models.

There are few examples of focused miscommunication analysis in the field of human-robot interaction. In Chapter 4 we presented an exploratory study of communicative errors related to the grounding model presented by Brennan and Hulteen (1995). Strategies for reducing miscommunication, like using back-channel responses were discussed by Trafton et al (2006). Breazeal et al (2005) analysed miscommunication in order to measure the effects of different non-verbal strategies that affect the efficiency and robustness of human-robot communication. Studies of miscommunication can also be performed on corpus data collected for other human-robot interaction purposes. One example is the work by Bug-

[1] A dialogue system partially inspired by the outcome of this work is described by Li (2007).

mann et al (2004) who investigated the amount erroneous or ambiguous statements for a specific task while collecting data to be used to construct a language model for route instruction for a mobile robot.

The study of miscommunication has attracted interest within the spoken dialogue community. There are several views of how miscommunication should be understood and approached. Martinovski and Traum (2003) provide an example of miscommunication analysis focusing on the identification of symptoms of miscommunication. There are also other more formal ways of classifying miscommunication, like the approach taken by Aberdeen and Ferro (2001) who classified miscommunication using four features: the type of error; surface evidence available to the user (like a repair act); the correction mechanism used (for instance, start over) and the outcome, whether the error was resolved or unresolved. Applied coherently this schema allows for using machine learning approaches to be used in the development process.

Approaches for evaluation used for classical spoken language user interfaces are hard to apply directly for human-robot interfaces. While the data used by Aberdeen and Ferro (2003) was only verbal, the multimodal character of human-robot communication complicates the discovery of error because non-verbal behaviour of users needs to be taken into account. More comprehensive evaluation models like PARADISE (Walker et al., 1997) or SASSI (Hone and Graham, 2000) are primarily focused on comparing different dialogue strategies and have been utilised to a lesser extent to evaluate human-robot communication since the focus in robotics has been on creating single instance systems, particular to a specific situation or scenario rather than comparing dialogue strategies. The complexity of setting up test scenarios has led to an interest in non-interactive methods, like the video-based evaluation, which was used to compare communicative behaviour described in Lohse et al (2008).

6.2 Miscommunication analysis in the design process

One of the steps taken to evaluate the dialogue model that was design for the Cogniron project was to analyse miscommunication. The analysis was targeted at integrating the analysis of miscommunication into the design process for a service robot. A schematic view of this is depicted in Figure 6.1. Incremental development of a natural language user interface is not straightforward. The idea of making up a

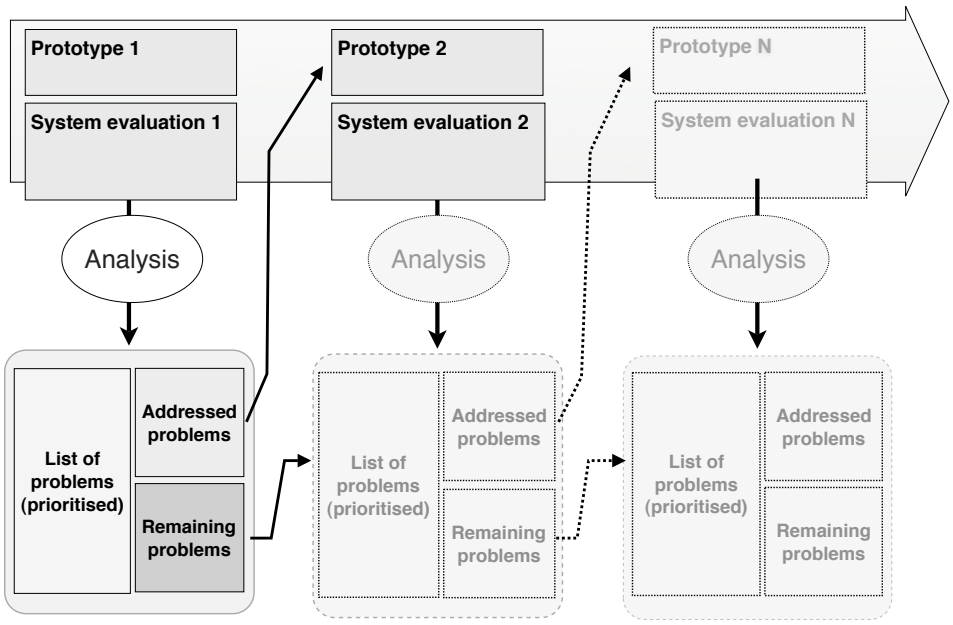


Figure 6.1 *A schematic view on how miscommunication analysis can be an integrated step of the design process. The analysis described in this chapter refers to evaluation and analysis of "Prototype 1". In the Cogniron project, a prototype that take design implications from the analysis into account has been built. This prototype refers to the "Prototype 2" in the figure.*

prioritised list of dialogue problems is an attempt to address miscommunication in a systematic way. The miscommunication analysis itself, identifying and describing what we came to refer to as trouble-spots is done without prioritisation. This means that this step could be performed as a distributed effort in a design team. Prioritisation between the different types of miscommunication is then a decision process that is dependent on factors like frequency, severity for the interaction, and to what extent this problem can be addressed with the current resources in terms of technology and know-how.

6.3 Analysis of the interactive sessions

The data collected for the analysis were transcribed video recordings from 11 interactive sessions which are part of the corpus described in Chapter 5. Table 6.1 contains general statistics for the sessions and the amount of utterances that occurred within a trouble-spot.

Table 6.1 *Statistics of the corpus material used in the miscommunication analysis.*

Session	Number of utterances	User/robot utterances	Utterances within a trouble spot (%)
02	131	71/60	18 (13.7 %)
03	180	106/74	10 (5.6 %)
04	65	39/26	1 (1.5 %)
05	176	93/83	14 (8.0 %)
06	167	86/81	5 (3.0 %)
07	139	85/54	2 (1.4 %)
08	108	61/47	4 (3.7 %)
09	132	66/66	12 (9.1 %)
10	152	78/74	12 (7.9 %)
11	128	78/50	1 (0.8 %)
12	148	80/68	3 (2.0 %)
Mean	**139**	**77/62**	**7 (4.77 %)**
Total	**1531**		

The way miscommunication analysis is used within the design process can be illustrated by the schema depicted in Figure 6.1. The result of an analysis focusing on miscommunication is a *list of problem types* (see Table 6.2). A frequency value is associated with each problem type as well as information about the state-of-the-art in terms of available technology and scientific status for the specific problem type. We can then analyse the list and assign priorities to each problem based on what is occurring frequently, is technically feasible and also scientifically interesting. Using this priority we can select a set of challenges that we can address when designing the next, and following versions, of our system. The way we approach this is similar to the approach known as *evolutionary design* (Alter, 2001).

In our analysis we adhere to the notion of miscommunication used by Traum and Dillenbourg (1996) which was introduced in Section 6.1, where miscommunication concerns the misalignment of mental states due to failures in producing or perceiving communicative acts by the interlocutors. The analysis was exploratory in order to identify an extensive number of problems of different types, frequency and severity. Using this approach we identified about 20 types of trouble-spots. We have deliberately defined a trouble-spot loosely to include cases that can be considered as causing minor problems. A trouble spot is a range of human and robot actions that covers some part of interaction that contains one or several cases of miscommunication.

We will limit our discussion in this section to the categories that were both frequent and more importantly, that could be attributed to problems related to the *dialogue model* that was evaluated in the use scenario. This means that problems that can be attributed to errors due to miscrecognition, own communication management (such as self-corrections) and general problems related to perception were noted, but not further discussed here since they do not contribute to the evaluation of the dialogue model. There were also instances of miscommunication occurring once, or very few times. Some of these had to do with lexical choices, for instance, the use of Swedish words rather than English or use of words that were deemed by the wizards to be out of the domain, such as emotive comments like *"good robot"*. Some problems were related to perception, meaning that the robot detected [2] that it could not perceive the user and therefore asked the user to *"get in front of the camera"*. These low-frequency problem categories are shown in the lower half of Table 6.2 (p. 128).

Types of miscommunication

In the corpus data we found sequences of containing miscommunication – trouble-spots, that were more frequent than others. The trouble spots are characterised as follows:

- *Mismatch:* Trouble-spots of this type are characterised as miscommunication and is caused by discrepancies between the user's model of system capability and actual system capability
- *Feedback-related:* These are trouble-spots that can be attributed to miscommunication can be categorised as:
  - Ill-timed feedback *causing incoherence* in the dialogue.
  - Ill-timed feedback *affecting the relevance* of the contribution.
  - Lack of feedback, meaning that the system *fails to respond* at an appropriate time.
- *Referencing:* Referencing, referring to the manner in which referencing of objects and locations was carried out.

[2] Based on the Wizard-operator's judgement about what is plausible given the scenario that is being enacted.

In the following we will examine these categories by providing and discussing a set of examples of miscommunication.

Mismatch

The miscommunication labelled "Mismatch" refers to cases where the actions of the users appear to be based on an understanding of the system that does not match the way the system is supposed to work, in other words there is a mismatch between the user's conceptual model of the system and the way it was intended to work by the designer (Norman, 1990). This category also covers what may be considered requests for tasks that are out of the domain, for instance, praising the robot by saying "Good work, robot". In the following example sequences which has been classified as mismatch has been underlined:

(124) **U:** stop robot
⟨ *robot stops its motion* ⟩
(126) **U:** turn[1] around
(127) **R:** Stopped following
(128) **R:** Cannot do that
(129) **U:** rotate[2]
(130) **R:** Cannot do that
(131) **U:** follow me

Example 6.3.1: Example of mismatch. The words '*turn*[1] and *rotate*[2] are not allowed according to the dialogue model.

In the exchange in the rows 124–131 in Example 6.3.1, there seems to be a mismatch between the robot's task capability and the tasks the user thinks the robot should handle. In this case, the user attempts to use a directive command to make the robot turn (underlined), something which the system does not support.

However, there are cases that are not clear cut, for instance when a user shows the robot an object by holding it in his hand instead of placing it on a flat surface. It is clear to the wizards that this gesture should not be handled by the system, according to the test setup, where the object recogniser was supposed to have this technical limitation. This limitation was also mentioned to user in the written instruction:

Table 6.2 *Trouble spots identified in the corpus material and the threshold level.*

Type of error	**#**
Feedback	13
Mismatch (communicative/task)	13
Reference	9
Own communication management	7
Signal non-understanding	6
Language	5
Lexical choice	3
Perception related	3
Restart	3
Topic shift	5
(others, single instances)	7
Total # Trouble-Spots/Utterances	**80/1531**

> *"You may use your hands to show a single object to the robot. Objects that the robot should know can be indicated if they lie on a flat surface like a coffee table. The surface needs to be free from other objects – the robot will use its vision system to collect information about the objects.*
> *Say the name of the object that the robot should learn to the robot and use your hand to point where it is."*

This type of problem can be said to belong in both categories, meaning that it is a communicative problem because the system fails to detect a gesture, but it is also a domain problem since the task lies outside the robot's capability. The mismatch in the sequence in Example 6.3.1 can be detected by the user as a rejection of the command "turn around" (in Example 6.3.1, row 126). Then the user chooses to adapt his command by using the synonym "Rotate" (Example 6.3.1, row 129). After the robot has responded negatively to the second turn command (in Example 6.3.1, row 130) the user resorts to using what we understand to be a fallback command namely "follow me" (Example 6.3.1, row 131). During the sessions it seemed that the users resort issuing to what they assume is a command that will work, when facing problems in the scenario.

Another problem is related to the users' erroneous inferences about the system's capability. Small objects, such as magazines, pens etc, were sometimes moved before being shown to the robot (as depicted in Figure 6.2). Users tried

to hold up objects in front of the camera. This was considered to be an error according to the task model which caused the communicator wizard to issue a repair. As the task was constructed in the experiment, all available objects were considered to be within the domain, meaning that the robot should be able to recognise them, the mismatch discussed here had to do with the manner in which the objects were shown to the (simulated) object detection. In a real system, it is likely that objects that cannot be recognised because they are out-of domain could give rise to similar miscommunication phenomena.

(132) **U:** this is a table
(133) **R:** Found dinner table
(134) **U:** this is a pen
⟨ *u holds up pen* ⟩
(136) **R:** Rearrange the objects please
(137) **U:** this is a pen
⟨ *u points to pen on table* ⟩
(139) **R:** Found one object

Example 6.3.2: Mismatch due to misconception regarding gesture recognition.

This behaviour is illustrated in Figure 6.2. In the example below, the user is holding the pen while uttering (Example 6.3.2, row 134). The repair (Example 6.3.2, row 136) then influences the user's actions and a pointing gesture is issued. The utterance "Rearrange the objects please" prompted users to adapt to the constraints of the system.

Feedback

In the data we noted several types of problems related to feedback. For an interactive system like a robot it is necessary, or even essential, to provide relevant and timely feedback to maintain an orderly and well managed dialogue (cf. Bernsen et al. 1998; Dybkjær et al. 1997; Klingspor et al. 1997).

(140) **U:** stop
⟨ *robot stops* ⟩
(142) **U:** this is a [table
(143) **R:** Stopped] following

Figure 6.2 *In the situation depicted in the top image, miscommunication occurs because the user is displaying an object, a magazine, in a way that the robot cannot understand (according to the constraints given by the dialogue model). In the lower image the user has adapted his behaviour to conform to the dialogue model. This is done by placing the magazine on a flat surface. In this case the user found a ledge on the wall where the magazine could be placed upright.*

Example 6.3.3: Incoherent feedback

We have identified problems related to timing, meaning that feedback is *ill-timed*, typically delayed, something which may render it incoherent, like in the utterances in Example 6.3.3, rows 140–143.

When the user utters "stop" (6.3.3, row 140) and then tries to specify an object (6.3.3, row 142) he is interrupted by the robot saying "stopped following" (6.3.3, row 143). Issuing "stopped following" (6.3.3, row 143) is thus non-relevant since the robot already stopped. At this point in dialogue this does not cause a severe miscommunication, but if the error occurs again, the user needs to adapt to the system's behaviour, something that might affect the attitude towards the system.

If we turn back for a moment to the sequence in Example 6.3.1 (rows 124–131), several phenomena that can be characterised as symptoms of miscommunication occur. Initially the user is stacking commands: first the user is commanding the robot to stop, and then he asks the robot to turn around. The response from the robot, meaning that it has stopped following the user (Example 6.3.1, row 127), in the contributions following the ones stacked by the user (Example 6.3.1 rows 124, 126) is delayed about four seconds.

The stacking is in itself not a sign of miscommunication, but the lack of feedback from the robot during the four seconds following the user's stop command can be regarded as an instance of the robot failing to make its contribution in a timely manner. This is based on the assumption that the dialogue model should explicitly provide feedback on each command. In this case the user seems to assume that change of topic, here understood as a task, is a communicative capability of the system. In human-human conversation, this type of stacking of topics is common, and we can therefore assume that the user is attributing human-like dialogue capabilities to the robot.

It is worth noting that the robot actually stops right after the user has given the stop command, well before issuing the response "Stopped following" (Example 6.3.1, row 127). This renders the utterance spurious and ill-timed. On the other hand, when the robot utters "Cannot do that" (Example 6.3.1, row 128), referring to the user's command "turn around" (Example 6.3.1, row 126), the user seems to

interpret this as relevant to the exchange and attempts another adapted version of the turn command (Example 6.3.1, row 128).

However, we should be aware that users seem to cope with ill-timed feedback as long as it does not conflict with the context and the task at hand. In following example the robot speech output overlaps[3]

(144) **U:** Stop
(145) **U:** this is a table
(146) **R:** Stopped following
(147) **R:** Found coffee table
(148) **U:** ahm
(149) **U:** please follow me

Example 6.3.4: Ill-timed feedback

(Example 6.3.4, rows 145–146) with the user's command. This seems to lead to the hesitation expressed by the user in the later utterance (Example 6.3.4, row 148). This is a symptom of miscommunication, but as a single instance it seems that it is not sufficient to cause a severe state of misalignment.

Another problem that was observed in the data was lack of feedback, meaning that the robot does not respond to the user's contribution before the user decides to make another contribution:

(150) **U:** this is a computer
{ *a few seconds passes*
(151) **U:** this is a computer
(152) **R:** Found one object

Example 6.3.5: Lack of feedback

In Example 6.3.5 (rows 150–151) as well as in Example 6.3.1 (rows 126–127) the robot does not respond to the user's input in a timely manner. In the study presented in this chapter we have not explicitly investigated the impact of timing. In Chapter 4 it was proposed that timing between utterances might be a factor to assess the acceptability of a speaking robot. Recently this has been verified by

[3]Brackets '[' and ']' indicate overlapping speech.

Shiwa et al (2008) who reported that the amount of time that users seem to accept is about two seconds, something which supports earlier findings by Miller (1968).

Reference

In the manner the system was simulated we allowed for a "robust" object recognition system, meaning that the system would recognise any object given that it was small enough and placed on a flat surface.

In the written instruction the users were told that the robot could find things it had been shown earlier during the session:

> *"Once the robot has been shown an object or a location it may be told to discover it again. The objects may be placed in another place than where the robot learned about it."*

This information may have contributed to miscommunication regarding reference. Although we added several pre-defined lexical entries for places and objects that we expected users to specify in the environment, there were some objects and locations missing. To refer to these places and objects in the simulated dialogue system we sometimes needed to give references with the words "object" or "location", because we did not have the time to enter new objects in the dialogue production tool. This caused some concerns for some users since they, as I have interpreted their behaviour, were not sure about the status of the recognised object. Other users did not seem to notice that the system was not giving an explicit reference to the object or location:

(153) **U:** this is a newspaper
(154) **R:** Found one object
(155) **U:** identify object
(156) **R:** Is this the object
(157) **U:** Yes
(158) **R:** Ok

Example 6.3.6: Unclear reference

In the example above the user attempts to get the system to name the object by saying "Identify object" (6.3.6, row 155). Once again it is hard to classify this trouble spot. This exchange could also be related to the category "Mismatch".

If we take on a strictly logical perspective, the positive response of the user to the utterance "Is this the object" cannot justify that the robot actually found the newspaper. The information can be said to have been negotiated to the extent that the users believed that the robot referenced the correct object. But since there is no pointing capability apart from the general direction indicated by the front robot and the on-board camera, there is no way of indicating precisely which object has been detected without actually providing verbatim or rephrased feedback.

Design implications

The miscommunication that was identified in the data collected during user sessions enacted using the Wizard-of-Oz technique was classified using twenty rough categories. These categories were further analysed with the goal of providing implications as a basis for a re-design of the robot's dialogue system. The design implications[4] were the following:

- Minimising mismatch. To reduce the mismatch between the users' conception of the system capability and the actual capability we need to familiarise new users with the functionality of the system. We suggest the introduction of an initial tutorial where the system explains its basic task and communication related capabilities. Another possibility would be to introduce context-sensitive help messages that can be displayed during the interaction.
- Use priming strategies to bias the user towards the use of words that the system can understand. This is based upon the understanding that *alignment* is taking place, meaning that the user and the system converges with respect to speaker style, lexical and prosodic structures (Pickering and Garrod, 2004).
- Provide relevant feedback using additional nonverbal (visual) feedback about the system state. For example, in the sequence in Example 6.3.1, rows 124–131), the (redundant) feedback "stopped following" is given too late (Example 6.3.1, row 127) and completely unnecessarily since the robot has already stopped, bringing the interaction out of synchronisation. In such cases, the execution of the task is a sufficient feedback signal. In this case the robot ac-

[4]The design implications of the system was developed based on the input from the design team at the University of Bielefeld and was presented in more detail in (Green et al., 2006b). I have briefly summarised them here to provide a complete picture of the analysis and re-design process.

tion should be considered as evidence of understanding (Clark and Schaefer, 1989).

- Provide more detailed feedback to resolve references, for instance, by equipping the robot with some kind of pointing device that gives a more detailed feedback than the direction of the pan-tilt camera. A a simple but efficient device would be to use a laser pointer mounted on top of the camera which will enable a very detailed reference. Additionally, verbal clarification can be initiated (like "Do you mean this object?").

The design implications regarding familiarisation were adapted and used to implement a tutorial by Li (2007). The challenge of giving relevant and detailed feedback still remain unsolved with respect to timing issues and reference as described above.

Discussion

In the study we found that miscommunication related to different aspects of communicative feedback seemed to cause problems. Another aspect that caused miscommunication had to do with the discrepancy between the users' beliefs about what the system can do and what the system actually can do, in terms of tasks and communicative capabilities. We have already assumed that miscommunication can be viewed as "a lack of alignment of agent's mental state" (Traum and Dillenbourg, 1996). In this case the users' expectations concerning what the system can do does not match the actual capacity of the system. Traum and Dillenbourg (1996) uses the notion of language as *action* (Austin, 1962) and argues that miscommunication can be viewed as instances of action failure. This is especially important in robotic systems. Any system action, both virtual and physical can have a communicative content. For instance, after issuing an action directive, like "go forward" the physical act of moving a small distance forward can be considered as communicative act which is performed under felicitous circumstances, in other words, a positive evidence of understanding (Clark and Schaefer, 1989). Likewise, the failure to act in the same situation can be viewed as a (dis-preferred) lack of action.

Another and perhaps more controversial aspect of miscommunication that was identified in the study concerns the places in the dialogue that were classified as a

trouble-spots due to observed lack of grounding between the robot and the user. In the example (below) the object "newspaper" remains ungrounded.

(159) **U:** This is a newspaper
(160) **R:** Found one object
(161) **U:** Identify object
(162) **R:** Is this the object?
(163) **U:** Yes
(164) **R:** Ok

Example 6.3.7: A dialogue that ends without explicitly grounding the object

The participant requests that the robot makes the reference to the object explicit by the request "Identify object". Then the robot asks, when turned towards the object, "is this the object?". The user then acknowledges this. The question is if the user interprets the robot's in-explicit question together with the deictic reference provided by the general direction of the robot platform as sufficient information.

6.4 Chapter summary

This chapter described how miscommunication analysis can be integrated in the design process for a communicative service robot. Using the corpus data of user studies with the Wizard-of-Oz technique a miscommunication analysis was performed. The miscommunication that occurred in the studied sessions could either be characterised as a mismatch between users' understanding of robot capability and real robot capability; problems related to ill-timed, irrelevant or lacking feedback; or issues related to the amount and quality of information needed for reliable referencing of objects and locations.

Using the miscommunication analysis as a basis, a set of design implications were created. To reduce or prevent miscommunication in human-robot communication it is important to:

- Familiarise users with the system to minimise the mismatch.
- Use strategies for priming users into use specific phrasing or style of communication, to reduce the complexity of language models.

- Provide relevant, well-timed and sufficiently detailed feedback using several modalities: speech, robot movements or communicative gestures from actuators, for instance camera-gestures or gestures displayed by an interface robot like the one designed for Cero (cf. Chapter 4).
- Feedback should be given in a timely manner.

The overall consequences for design of human-robot communication can be summarised as follows:

- Miscommunication should be designed for, since it provides an opportunity for learning system boundaries. This adheres to the view of miscommunication taken by Martinovski and Traum (2003).
- Grounding status of objects and locations needs to be represented in the system. Task-related information available in the system should be possible to access, to inform the user of what the system has learned, for instance, objects and locations that have been specified.
- Robot actions may have a communicative function, something which means that consequences of robot actions need to be taken into consideration when generating communicative actions.

CHAPTER 7

Design Implications for Information on Communicative Status

As noted in the previous chapter some miscommunication can be attributed to the inability of robots to take actions in time or to provide timely and relevant feedback. The purpose of this chapter is to discuss how feedback related to the robot's communicative status affects the quality of communication, and how this can be taken into account in the design of natural language user interfaces for communicative robots. This will be done in two steps, first by discussing corpus observations of miscommunication related to the communicative status of the robot and the user. Second, design implications regarding perception feedback will be discussed and motivated, based on the analysis of corpus data and examples of design found in literature.

7.1 Initial observations: ill-timed or lacking feedback

I will start by pointing out some challenges to communicative quality based on experiences from the work with the two systems introduced in the previous chapters.

In the example[1] below, which was collected in the work with the Cero system (Chapter 4), the robot responds only after several utterances by the user. There is no feedback given by the robot that gives the user any guidance of what the robot perceives, meaning that no communicative contact is established until the robot provides positive task-feedback (Example 7.1.1).

[1] The whole example is also given on page 85

(165)	**U:**	Go to Lars office!	//5 sec pause//
(166)	**U:**	Go to Lars office!	//5 sec pause//
(167)	**U:**	Cero, go to Lars office!	//2 sec pause//
(168)	**R:**	Go to Lars office?	

Example 7.1.1: Lacking contact and perception feedback

In the Wizard-of-Oz data collected to evaluate the dialogue design for Cogniron we experienced situations, like the one in the example below (Example 7.1.2), where miscommunication occurred due to what could be attributed to some kind of misperception of the "system". Even if the wizard operators attempted to achieve rapid and relevant responses there were still cases where slow responses caused feedback to be ill-timed or lacking:

(169)	**U:**	Please follow me	
(170)	**R:**	Robot is following	
		{ *about 10 seconds passes* }	
(171)	**U:**	This...	
(172)	**R:**	Stopped following	*ill-timed feedback*
(173)	**U:**	...is a switch	
(174)	**R:**	What did you say	*request-repair/SNU*
		⟨ *Test leader encourages participant to try again* ⟩	
(176)	**U:**	Eh, repeat please	*req-repair*
(177)	**R:**	What did you say	*request-repair/SNU*
(178)	**U:**	What, eh please follow me	
(179)	**R:**	Robot is following	

Example 7.1.2: Miscommunication: ill-timed feedback, signaling non-understanding (SNU) and request for repair (req-repair)

To some extent this could be an effect of the self-imposed constraints on the wizards to simulate a realistic system. Another reason was that wizards could not react in time to provide feedback, even if they had wanted to, partly due to technical limitations and partly because they need to react to and plan what to say before using the wizard tool. The general observation is that slow or lacking feedback

from the robot seems to lead to frustration among the participants and a sense of being disconnected from the system. This made me reflect on the possibilities of providing feedback on communicative status automatically.

As the technical implementation of such a system falls out of scope of this work I will focus on motivating why giving such feedback is desired in human-robot communication. This discussion will be based on theories of human-human communication and by analysis of examples from the corpus.

7.2 Perspectives on feedback

To understand what role feedback on the communicative status of the robot may play in human-robot communication we can turn to theories and accounts of feedback in human-human communication. I will initially introduce some relevant concepts from research on human-human communication.

Contact and perception feedback

In Chapter 2, the notion of communicative feedback was introduced as one of the main ways in human-human conversation to create common ground between interlocutors. Most accounts of communicative feedback concern understanding and attitudes to the main evocative content of the speaker, which in a robotics context could a task specification, like "go to the kitchen" or "this is an orange". But feedback related to the communicative and perceptual status and the willingness or ability engage in communication are also crucial in order to establish whether feedback concerning the main content can be perceived and understood. Feedback concerning the perceptual state and the willingness to interact provide means for the interlocutors to display their own communicative status and to evaluate the communicative status of others (Allwood et al., 1991; Brennan and Hulteen, 1995; Bunt, 1999; Clark and Marshall, 1981). Perception feedback consists of communicative actions that report on the evaluation of the perceptual state of the individual and their ability to perceive others, like "I hear" or "what did you say?" (Allwood et al., 1991). Feedback can be expressed verbally as well as multimodally, through gestures and body moves (Allwood, 2002; Gill et al., 1999). The constructed example below (Example 7.2.1) is intended to illustrate various types of providing feedback.

SP is a salesperson. Customer C is shopping for hats:

(180)	**C:**	⟨ enters the store looks at A⟩	
(181)	**SP:**	⟨looks at B⟩	+FB, contact
(182)	**C:**	I would like to have that one.	
(183)	**SP:**	What did you say?	−FB, perception (hearing)
(184)	**C:**	That one ⟨points⟩	
(185)	**SP:**	⟨shrug⟩	−FB, perception (reference)
(186)	**C:**	**That** one ⟨points⟩	
(187)	**C:**	⟨nod⟩⟨gaze on object⟩	+FB, reference
(188)	**SP:**	Here you go!	

Example 7.2.1: Feedback related to perception and willingness to interact. '−FB'= negative feedback, '+FB'=positive feedback. Hearing refers to the user's ability to perceive verbal utterances through the audio channel. Reference refers to feedback related to deixis.

Attention drawing

Exchange of feedback regarding the communicative status plays an important role in the management of how a (partially) shared environment can become mutually perceived by participants in conversation. One important factor for this relies on the circumstance that perceptual attention of humans can be *drawn* towards features in the environment and in-between participants in conversation. To explain this different attempts have been made to provide characteristics for stimuli that has the ability to draw our attention. It is argued that "transients", meaning sudden shifts in light (Pashler et al., 2001) has the ability to draw the attention of humans. Clark (1996) uses the term "perceptually salient" to describe these phenomena. Visual and auditive stimuli are said to have the potential of attracting attention, and stimuli with a high relative strength of this potential (salience) are most likely to draw attention (Pashler et al., 2001). It is well known that cognitive processes allow humans to focus their perceptual attention actively (Cherry, 1953). Salience therefore seems to be determined by the individual's mental state as well as the type and relative strength of the stimuli occurring in the environment.

In the context of human-human communication some types of stimuli seem more relevant than others. Human speech is a stimuli with a high relative sali-

ence (Bregman, 1994). Another strong stimuli is human gaze. Gaze behaviour seems to be one of the primary keys for finding cues for human visual attention. Humans have a good discrimination of the line of gaze of others (Gibson and Pick, 1963) and most psychologists generally agree that humans have modular perceptual subsystems for recognising gaze directions (Langton et al., 2000; Wilson et al., 2000). The direction in which another person is looking therefore provides participants in conversation with important cues to the focus of attention of each other.

Feedback to create perceptual co-presence

To be able to communicate in a coherent way about their environment, participants in conversation need to ensure that they share the same view of the surrounding physical context. This is referred to as having a *shared perceptual base* or *being perceptually co-present*. To be aware of the same perceptually salient features in the environment and to become aware of how these features are perceived by others is an important part of the communication process. This is stressed both by Goodwin (2000) and Clark (1996; 2004) who argue that this is a continuous process which takes several modalities into account. Clark also stresses that sharing the same perceptual context does not immediately lead to a perceptual co-presence, there also needs to be some salient event that leads to participants focusing their mutual attention to the same thing. In such a state the salient perceptual events have been grounded, something which allow participants to infer the meaning of utterances that are related to these percepts. A salient event is a perceivable event of some kind that draws the attention of the participants, like in the following constructed example (Example 7.2.2).

A and B are walking in a field. Lightning flashes and thunder is heard.

(189) **B:** That was close. Better take cover.

(190) **A:** Yes

Example 7.2.2: Gaining shared perceptual bases through a perceptually salient event.

Indication of perceptive behaviour on the part of one dialogue participant may be interpreted as *eliciting* contact feedback by the other. One example of this is when people attempting to engage in communication by gazing at each other.

A and B are playing chess. Both are looking at the chess board:

$\langle$ *B looks up at A* $\rangle$

$\langle$ *A looks up at B* $\rangle$

(193) **B:** Your move!

Example 7.2.3: Establishing mutual co-presence. A and B are mutually perceiving each other.

The information necessary to establish perceptual co-presence may come from gaze as well as from other means of communication, like speech. Clark (1996) lists three main ways to achieve mutual perceptual co-presence:

- Activities that indicate perceptual processes.
- Gestural indication relative the environment, for instance, deictic gestures like gaze, posture or hand gestures.
- Salient perceptual events, meaning events that indicate significant change of the physical environment like a beep from a telephone or a flickering light.

Another aspect of perception feedback concerns its ability to draw attention to or change the perspective of the physical and informational context. Referencing and asserting qualities to the physical and informational context is an integral part of situated activity. According to Bunt (1994) conversational acts may have effects on several levels, including the perception of the physical context. Goodwin (2000) describes how feedback is used in the process of configuring context, meaning that it is used to acknowledge change or influence the meaning of what is perceived by participants in conversation, for instance by turning gaze towards objects referenced in conversation.

7.3 Corpus observations

Up until now I have discussed some theoretical notions of feedback in human-human conversation. The first goal of this chapter is discuss how feedback on the communicative status of the robot affect the quality of interaction. I will approach this by discussing corpus analyses of sequences of interaction where inadequate or lacking feedback on the robots perceptual status seems to lead to miscommunication. This study focused on the following aspects of the interaction:

- How utterances and body language are used by participants in their attempts to evoke reactions of the robot.
- How the participants actively monitor the actions of the robot.
- The manner in which the environment is used by the participants to ground references to objects and locations.

The analysis of the videos taken from the corpus (described in Chapter 5) were carried out using a standard media viewer. While browsing the trial sessions I looked for scenes where miscommunication related to perception occurred. The time of these scenes were noted down on paper. Then each scene that was found interesting was studied in more detail, taking the aspects listed above into consideration. The focus of the analysis was to comprehend as many aspects of the situation that could affect the interaction as possible, with respect to communicative status. Special consideration is taken to what type of cues, or salient events, that occurred in the environment or on the robot prior to any action of the participant. This could be verbal utterances from the robot, robot body movements, movements from the attached camera, robot noise or any other event that could be perceived by the participant. I also paid attention to where the participant was looking, by estimating gaze directions. Figure 7.1 shows an example where gaze is important to understand what is going on in the interaction sequence.

Corpus examples

In the following I will discuss some examples taken from the corpus (described in Chapter 5). These examples reflect the way the participants in the study tried to evoke communicative action of the robot in order to establish communicative contact or take action aimed to assess to what extent the robot is perceiving its user. The behaviour and actions taken by the participant to assess the communicative status of the robot are perhaps best understood as steps that lead toward the overarching goal in the scenario: to teach the robot new objects.

Monitoring

The example in Figure 7.1 gives an example of the way most participants continuously monitored the robot during the interaction. In the interaction sequence the participant has commanded the robot to follow. The robot slowly starts to

Figure 7.1 *An example of continuous monitoring of robot activities taken from the corpus. The top left image shows the participant looking over the shoulder, waiting for the robot to start moving. The lower left image shows the participant looking towards an object (the TV set). The image to the right shows the participant turned towards the robot, looking at it as it comes closer to the participant. The next action taken by the participant (not shown here) is to name and reference the TV-set.*

move towards the participant who is glancing over the shoulder while moving forward. This is similar to what Schegloff (1998) observed using the notion of "Body torque", which describes a situation where the direction of the lower body relative to the direction of the upper body "projects change" into a new situation. In this example the body of the participant is turned towards the goal (the TV-set) and the upper body and the gaze are turned towards the robot. When the robot eventually moves, the participant walks towards the TV-set with her back turned. This can be interpreted as if the participant has found sufficient evidence of understanding and interprets the initial robot movement as a commitment, on the part of the robot, to follow the participant. As the distance between the participant and the robot has increased, and the wall beside the TV-set is limiting the possibilities for the participant to move any further, the participant turns towards the robot. The robot is then monitored closely while it moves towards the position of the participant close to

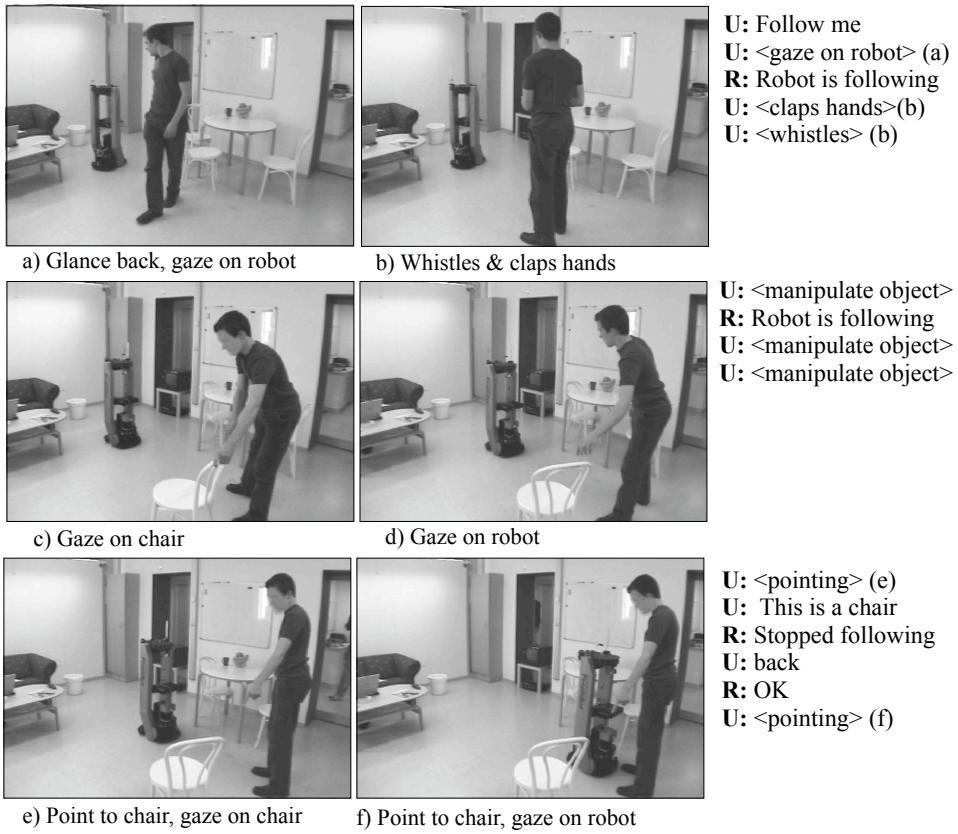


Figure 7.2 *An example showing a sequence where the robot fails to respond to the participant's requests for perceptual and contact feedback.*

the TV-set. During the monitoring sequence the participant is walking backwards and keeps looking at the robot. Once the participant reaches the wall behind her, she looks briefly at the TV-set, and then looks back at the robot while pointing to the object and asserting a name to it.

Attempting to establish contact

The example in Figure 7.2 is also taken from the corpus and shows a participant that is about to show a chair to the robot. In Figure 7.3 (p. 149) the same scene is displayed as tracks similar to a musical score annotation, showing multiple activities in several modalities in parallel.

In the scene the participant commands the robot to follow him. The next thing that happens is that the robot acknowledges the follow command, and then the following behaviour is started. The robot then starts to move and turns towards the participant. The participant, however moves slightly too quick, and the robot cannot follow him (Figure 7.2:a). Once the participant notices this, he attempts to attract the attention of the robot, first by clapping his hands and then by whistling (Figure 7.2:b).

It is reasonable that the participant is performing this in order to get the robot to continue its following behaviour, assuming that once contact is established, the following would continue. If we analyse this in terms of grounding, the participant first attempts to establish contact by securing an open communication channel, before any attempt to solve a task can be made, like specifying a task. Establishing contact then becomes fundamental for being able to get the message through. To establish contact, the participant needs to take the perceptual status of the robot into account.

The next thing that happens in the interaction sequence is that the participant walks back in front of the robot with what appears intended to re-establish the perceptual attention of the robot (this sequence occurs between images b and c in Figure 7.2). As the robot then starts to move and reports this by saying *"Robot is following"* the participant moves so that he stands in the middle of the room (Figure 7.2:c-f). On the way the participant manipulates a set of objects, presumably to make way for himself and the robot (Figure 7.2:c-d).

In what follows it seems that the user is configuring the environment so that he can provide a deictic reference in a position that is comfortable both to him and with respect to the way the participant think the robot's object recognition system works. The configuration of the environment seems to support the next step in the overall activity, to reference and name an object to the robot. During the sequence that follows the participant also glances back and forth between the chair and the robot as the robot approaches (Figure 7.2:c-d).

When the robot has come close the user commands the robot to "stop" (Figure 7.2:e) and then to go "back". This is done while pointing at the chair. My interpretation of this is that the participant is unsure whether the robot is actually perceiving the pointing gesture issued in (Figure 7.2:f).

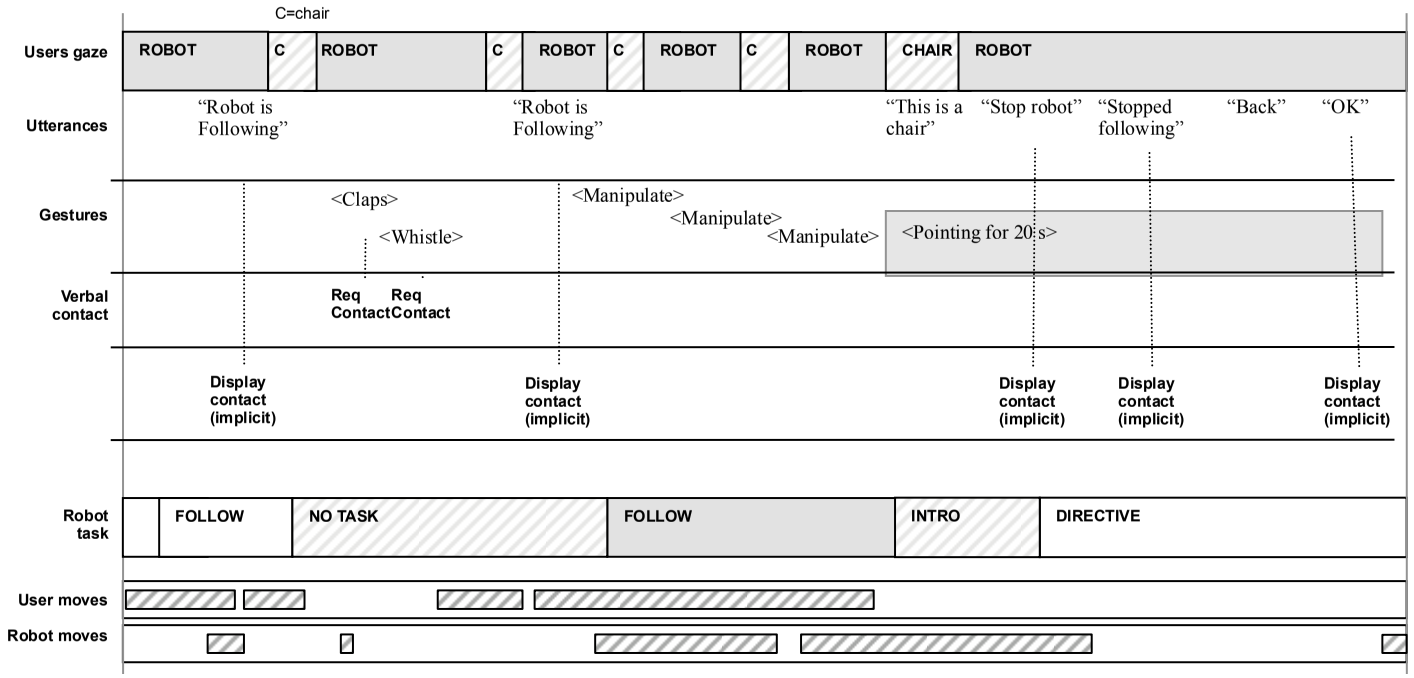


Figure 7.3 *A track analysis of the example depicted in Figure 7.2.*

Figure 7.4 *The participant is seeking gaze while providing a combined verbal and gestured utterance intended to turn the robot.*

Attempting to establish contact by seeking gaze

Another example which exemplifies participants looks for feedback from the robot in order to establish that a command has been received is shown in Figure 7.4. The verbal actions below are very few, instead it is the active behavior to establish gaze contact with the robot which indicates that some kind of feedback is necessary to alleviate the miscommunication that occurs:

(194) **U:** Follow me robot
(195) **R:** Robot is following
(196) **U:** ⟨TURN GESTURE⟩
⟨ *Stoop in front of robot* ⟩
(198) **R:** Please stand in front of the camera
⟨ *Participant gets back in front of camera* ⟩
(200) **R:** Robot is following

Example 7.3.1: Seeking gaze contact (Figure 7.4)

In the scenario the participant first commands the robot to follow. As the user moves away too quickly the robot does not start moving. The participant then stands beside the robot and leans in front of it (as is seen in the middle right hand picture) and tries to make the robot turn by gesturing while actively seeking the gaze of the robot. In this situation the outcome was that the robot did not turn, because the (enacted) constraints of the gesture recognition given in the scenario did not allow this type of gesture. Then when the robot provides a directive command, prompting spatial actions (asking the user to "get in front of the camera"), the situation is resolved and the following behaviour can recommence.

7.4 Information of communicative status on different levels

In order to design a robot that can provide feedback based on information regarding perceptional information available in the system, we need to consider a design space that ranges from communicative feedback using natural language, physical action and design which is intended to trigger affordances that indicate perceptual activity. The review of the different perspectives on human-human communication in this chapter and in Chapter 2 provides some of the necessary background that allow us to consider how feedback behaviour can be adapted to human-robot com-

munication. The terminology used in this section provides a challenge as feedback is generally considered to be a reaction to some initiating input. We could then try to distinguish between feedback, given as a response and information about communicative status, which can be given at all times. But for practical reasons I will refrain from making a categorical distinction between low-level feedback and continuous information about communicative status of the robot. In this section I will discuss some general considerations for providing feedback using an embodied robot, before turning to the discussion of design implications in the next section.

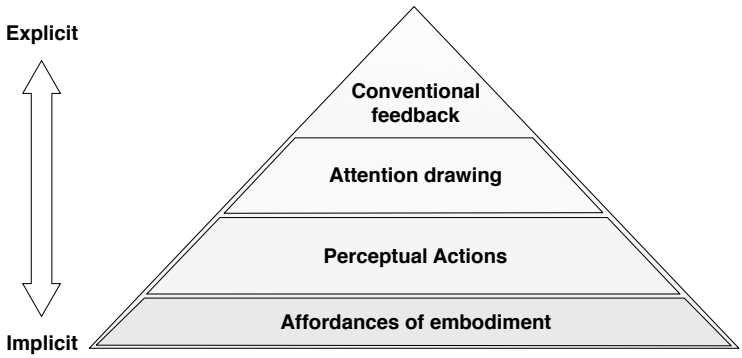


Figure 7.5 *Feedback can be given on different levels. Conventionalised feedback refers to feedback based on system information on a symbolical level, normally verbal and gestured feedback provided by the dialogue system. Affordances of embodiment provide information that is relevant to the interpretation of feedback.*

Information regarding the communicative state of the robot can be displayed on what can be understood as different levels of conventionality and arbitrariness as indicated in Figure 7.5. The shape of the figure is intended to visualise to what extent a specific type of feedback is present in a system, depending on to what degree it is conventionalised and explicit expressed. For instance, conventional feedback is provided deliberately but is not present at all times. Affordances of embodiment are always present, and may be interpreted as feedback in certain situations. For instance, when the robot is turned towards the participant, it can be interpreted as willingness to interact. A set of examples are listed in Table 7.1.

Level	Examples
Conventional	OK! + ⟨nod⟩ ⟨nod and turn body to object⟩ ⟨start task action: moving forward⟩
Attention drawing	⟨turn towards user⟩ ⟨make a loud beep⟩
Perceptual actions	⟨turn camera towards object⟩ ⟨sound from ultrasonic sensors⟩
Robot embodiment	a direction provided by the direction of movement noise from on-board computers and motors

Table 7.1 *Examples of feedback on different levels*

In dialogue systems feedback to users can be given explicitly by displaying conventionalised verbal and gestured signs ("Conventional feedback" in the figure). We can relate this to the model introduced by Brennan and Hulteen (1995). In their model, they considered what can be viewed as explicit feedback, that could be given on eight different levels, corresponding to different depths of grounding.

Feedback can also be less conventionalised, trying to draw the attention to the system, possibly to make the users aware of explicit feedback to be given in the immediate future ("Attention drawing"). This feedback may consist of audio or visual cues, such as sounds, conversational fillers (Shiwa et al., 2008), or gestures, like the head-nods proposed in Chapter 4.

As the robot's affordances and physical actions taken to perceive its environment are non-arbitrary, meaning that they are taken as part of the system's way of operating, they can lead to miscommunication. To some extent this explains some of the miscommunication that could be observed in the corpus material, for instance, when the user seeks the gaze of the camera. This may lead to false affordances (Gaver, 1991). An example of this is when the camera appears to display a gaze behaviour, something which is misinterpreted as robot is "seeing" in this direction.

7.5 Design implications

The second goal of this chapter concerns how feedback (or information) on the communicative status of the robot can be taken into account in the design of natural language user interfaces. In this section I will address this by introducing a set of design implications regarding feedback using the observations made in the corpus material and the theories of human-human communication as a backdrop to the discussion. The design implications that are being introduced in the following should be seen as ways to actively consider how feedback about communicative status can be made part of system design. This can be feedback which can be seen as a type of side effect from the robot behaviours, either actions taken by the system to solve tasks or actions related to robot perception. At the other end of the spectrum we find feedback that can be given deliberately to indicate communicative status, by verbal or gestural means.

Robot embodiment indicates perceptual activity

I will diverge slightly from the understanding of feedback as reactions on user actions and claim that the physical embodiment, through its affordances provides information that is as relevant for communication as feedback. In Figure 7.5 this is represented as a base layer of information ("Affordances of embodiment").

The concept of *affordances* (Gibson, 1979) perhaps needs some further explanation. In a psychological perspective objects afford different kinds of behaviour, for instance, a door handle affords pulling, a button affords pushing, which are determined by cognitive and cultural factors. Affordances of the human body, especially the human face, play a special role during language perception and understanding. For instance, facial gestures (like mouth movements) seem to aid the auditory perception process (Massaro, 1998). The affordances of human-like embodied agents are believed to provide strong cues for interaction and people seem to find them engaging (Cassell et al., 2001).

The robot's physical embodiment, such as the presence of a face, eyes or other design elements, can be interpreted as indicating perceptual activity of the system. Physical actions that are performed by a service robot may also be interpreted in a communicative manner. Examples of such activities can be actions performed by the robot when it actively tries to perceive its environment, moving a camera,

turning the robot body or firing ultrasonic sensors. In this manner a robot can display perceptual activities more or less overtly, ranging from a state where no perceptual actions are visible to the user to a state where the user may interpret system actions as perceptual activity:

- Display of robotic perceptual activity, for instance sound from ultra sonic sensors, sound from motors when moving or camera movements.
- Display of bio-mimetic perception activity, like the looking behaviour of an artificial creature.

Actions that are related to the task the robot is solving may also be interpreted as feedback, for instance by understanding compliance to a request to move as evidence of understanding (Clark and Schaefer, 1989). The act of moving can then be interpreted on several levels, both as something factual, where the act of moving, in order to accomplish the task, can be seen as feedback on the positive attitude towards the main evocative intention[2] of the user (to get the robot to move), and the display of the system being in a state where it is able and willing to understand further input (continued contact).

Continuous indication of availability

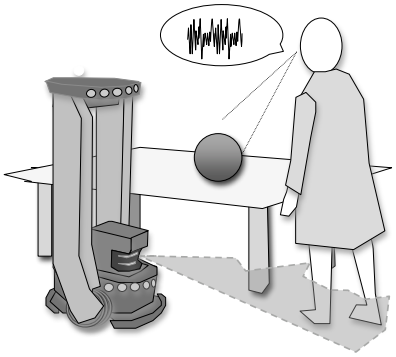
The observations in the corpus that the participants were closely monitoring the behaviour of the robot and also seemed to react on what appeared as communicative cues indicates that users want to know if the system is available. Seen as a design implication, this means that the system should continuously indicate to what extent it is available for communication. This can be understood as the system should provide feedback concerning its willingness to interact and its ability to perceive the users' input. These design implications can be phrased as:

- Provide continuous indication of perceptual availability and display feedback related to the willingness to interact.

In order to display such feedback it is necessary that the system can take both the perceptual status of itself and of the user into consideration, the latter to estimate to what extent the robot is successful in communicating its availability to the user. It should be noted that even humans have to rely on an estimation of to what extent

[2]This is explained in section 2.2, p. 20.

conversation partners are perceptually available. In a robot this can be modelled explicitly, for instance by displaying feedback based on, for instance, a change in system state that indicates that a user has been detected in the vicinity of the robot and therefore can be assumed to be perceptually present. By providing positive feedback in this situation, the robot invites further interaction by the user (who knows that the robot is paying attention). Positive feedback that is displayed based on the detection of a perceptually available user is then a way of communicating that the system is willing to interact.



⟨*robot is positioned in front of a table together with the user*⟩
R$_1$: I sense that you are here^A Show me something, please!
U$_2$: Okay, this is a ball ⟨user points⟩
⟨a couple of seconds passes⟩
R$_3$: You are still around^B , do you want to show my something more?
⟨ *robots turns towards table*⟩

Figure 7.6 *Cues available to the system – detected speech, gaze, objects identified and user presence (for instance, by using laser range finder), together with a possible dialogue.*

There are several different possibilities for detection of human presence. The point in this discussion is not to give an extensive survey, but at least one is needed to motivate the design proposed below. Using a laser rang finder to detect pairs of legs is a method that has been successfully used for finding and continuously tracking humans, something which enable the design of following behaviours, has been explored by (Fritsch et al., 2003; Topp and Christensen, 2005). Using this type of leg tracking it is possible to estimate where people are, but not where they are looking.

Figure 7.6 shows a synthetic dialogue which is based on the information, available to the system, that the user's legs have detected. The underlying idea, in its simplest form, is that only the cues of human presence is required in order to

provide positive feedback in Figure 7.6. By assuming that the current service[3] provided by the system is to allow the user to teach new objects to the robot (as in the Home Tour scenario), the robot can prompt the user to show it something, once it detects that the user is still present. It continuously displays its communicative availability and its willingness to interact.

Displaying perceptual availability

Several ways of continuously providing information on the communicative status of the system have been proposed in the literature on human-robot communication, and I will discuss a set of relevant designs. One way of displaying continuous availability can be provided by showing idling behaviours, for instance, moving the robot actuators, like, arms, legs, head or camera (Böhme et al., 2003; Yamaoka et al., 2006). Another way of indicating communicative availability is to maintain eye-contact, by moving head and camera. Yamaoka (2006) also describes how physical contact can be established by touching the hand of the conversational partner. The robot also reacts to touch by the approaching conversational partner. Idling behaviours provide the opportunity for interaction partners to make a judgement about the possibility of engaging in interaction because the robot appears to be switched on and possibly available for communication.

One of the prerequisites for continuously providing information about the communicative status is that the system is able to refrain from making utterances at a stage in the communication process when the interaction partner has difficulties in perceiving them. One interesting approach for doing this is proposed by Martinson and Brock (2006; 2007) who describe a system to support auditory perspective taking, meaning that the robot can adapt its communicative behaviour to the sound-scape of the surrounding environment and the position of humans present. To some extent this means adapting to environment noise, by changing the volume of speech output. The system also considers the impact of the noise in the surrounding context in order to adapt to the perceptual status of the interlocutor, by pausing for interruptions that come from contextual noise or sound sources in the environment. This also means that the system can alert the interlocutor to the fact that it is about to continue speaking by saying, “As I was saying...”.

[3]The Home Tour was limited to identifying an *object* placed on a flat surface. The robot had no means of identifying human presence.

Grab attention to be relevant

In the corpus data it could be observed that the participants were continually monitoring the robot while it was moving. When the robot was standing still, the participants focused on preparing the next steps in the interaction, for instance by moving objects, or they were reading on the instruction sheet. Arguably we can also assume that if a robot has been used for a longer period of time and the novelty effect[4] wears off, the need for monitoring its behaviour will be reduced. For cases when a user is not paying attention to the robot, the robot may need to grab the attention of the user. This is to assure that some particular feedback or information is given at a specific point in the unfolding interaction so that it is relevant for a particular communicative purpose. The implication for design can be phrased as follows:

- Grab attention using perceptually salient cues.
- Provide feedback when attention is established.

For example, camera behavior can be used to actively grab attention, a way of actively making contact. We have noted in our corpus material that gaze behaviour of the camera is interpreted in this way.

The ways gaze attention is used in human-human conversation has been investigated by Goodwin (1981) who argues that it is important for the speaker to have secured the gaze of the listener in order to produce a coherent sentence. This has been formulated as a conversational rule: *"A speaker should obtain the gaze of his recipient during the course of a turn at talk"*. Goodwin (1981) also provides an account for how to get the hearer to turn gaze towards the listener: either make a restart by uttering a phrase in the beginning of the utterance and then provide a coherent phrase. Another possibility is to pause slightly in the middle of the sentence. At the point of the pause/restart the gaze of the hearer is turned to the speaker.

Gaze is important for getting the attention of the communicative partner also in human-robot interaction. Design of interactive robots that involve gaze behaviour is described in different ways in the literature. For instance, gaze was used to estimate engagement by Sidner et al (2004). In their study they found that a robot that captures the attention using head movements and gaze direction is more engaging: the participants gazed back, engaged in mutual gaze and responded to the robot's

[4]The novelty effect has been observed by several authors (Gockley et al., 2005; Kanda et al., 2004; Salter et al., 2006)

commands more than a robot that only talked. In another study on robotic gaze, Martinson and Brock (2006; 2007) describe how the robot turns towards the face of the interaction partner, primarily in order to increase the intelligibility of spoken utterances, but also because it allows the user to move around while interacting with the robot.

In my view a robot provides and grabs attention continuously when it turns towards the user. The behaviour of turning towards the user as an attempt to attract and maintain attention has been described by Haasch et al (2004). In addition to turning towards the user Holzapfel et al (2006) describe a way of attracting attention by issuing spatial prompts (communicative actions that incite spatial action) to engage the user in interaction, by saying "Hello, please come closer!" and then "use the headset to say hello" to establish verbal communicative contact.

7.6 Means of displaying communicative status

The basic assumption underlying this work is that miscommunication can be reduced if feedback regarding the communicative status can be provided by the robot. In the previous sections we have seen how miscommunication caused by lacking feedback of the communicative state of the robot occurred in many of the use situations of the corpus. The implications for the design of human-robot communication with respect to feedback on the communicative status of the robot provide us with a situation where we need to consider the communicative status of *both* the user and the robot.

In the examples in Figures 7.1 and 7.2 the user and the robot get information from each other using their perception, the surrounding context and their background knowledge. The question we need to address is what type of information about perceptual status we need to represent to create a system which can provide feedback of its communicative status? In the examples the robot provides feedback/information on communicative status on different levels which can be perceived by the user. Information about communicative status can be given by:

– spoken utterances, together with knowledge of what one could expect from someone (or something) that can use linguistic structures like, *"I am following",*

- camera gestures together with knowledge (a "mental model") of what a robot mounted camera is able to capture, and,

- robot movement and spatial position that indicate that the robot is focusing its attention on a specific task.

Work on multi-modal anchoring for autonomous service robots (Fritsch et al., 2003) may serve as examples of use of how information cues for keeping track of people in the environment is useful for producing feedback. In an attempt to address the full-scale problem of natural human-like human-robot interaction, Scheutz et al (2007) have built a robot which uses many of the cues and information parameters similar to the ones proposed in this section. But even if the goal is to achieve human-like interaction, there are challenges regarding the monitoring of perceptually salient activities. Scheutz et al (2007) reports that their robot mistakenly takes *"the lack of intelligible output from [the speech recogniser] as silence (and disinterest) on the part of the person."*. It seems that failing to recognise that the user attempts to establish contact leads to the failure of the system as a whole to respond in an appropriate way.

7.7 Chapter summary

Continuous monitoring of the communicative status of other conversation participants is an integral part of the communication process between humans. Linguists and psychologists stress the importance of feedback in the grounding process to evaluate the communicative status of participants in conversation. To know whether the interlocutor is perceptually available, able to seem, listen and react to perceptually salient events is important to enable communication focused on achieving joint tasks.

Feedback and information on the communicative status can be given on different levels, ranging from conventional feedback by displaying verbal and gestured utterances, signals intended to draw attention, physical actions that signal perceptual activity to designed features that provide affordances to indicate perceptual capability of the system.

The observations in the corpus of human-robot interaction indicated that the participants continuously monitored the behaviour of the robot and were actively

seeking the perceptual attention of the robot. This indicated that they were lacking information about the communicative status of the robot.

The theoretical understanding of the relevance of perception and contact feedback together with the observations made in the corpus, give rise to some design implications for how to provide feedback on the communicative status of service robots.

To produce feedback/information on the communicative status we can provide feedback on different levels, for instance, conventional feedback: verbal utterances, gestures etc. We can also display communicative status by drawing attention using audio, visual or spatial actions. It is also possible to take the view that feedback and information of the communicative status can be given through perceptual behaviour like camera movements together with affordances provided by the robot embodiment, like a heading or direction of movement. Another dimension is the task configuration of the robot, its movements and pose which can be used to infer whether the robots' activity can be regarded as feedback given to the user.

The phenomena concerning perceptual behaviour observed in the corpus, namely that participants continuously monitor robot activities and attempt to establish contact support the following design implications regarding feedback on the communicative status:

- Grab attention using perceptually salient cues.
- Provide feedback when attention is established.
- Provide continuous indication about the system's perceptual availability.

The implementation of a system that represents this information falls out of the scope of this thesis. We have seen a shift in focus in research on communicative robots in the last decade, from robots with a single interface modality, like speech input, to multi-modal interfaces where sensor data is fused. This development calls for approaches that can take the perceptual status of robots and users into account.

CHAPTER 8

Spatial Influence as a Design Element

This chapter is focused on the way a robot can actively influence the spatial behaviour of the user. In this chapter I will introduce and discuss the concept of *spatial prompting*, and more specifically try to answer the research questions whether spatial prompting can be motivated empirically and how this can be used in the design of communicative robots.

The understanding of space has been studied in depth, for instance in social anthropology, and the term *spatial prompt* has been used in relation to the discussion of space syntax (Hillier and Hanson, 1984) and territoriality (Sack, 1986). Widlock et al (1999) uses the term "spatial prompt" to describe how a specific feature of a building, an *olupale*[1], projects change in social behaviour. In the following the term is used to capture phenomena that are related to actions that the robot can take to influence the behaviour of people. Phenomena related to spatial influence have been studied by Lewin (1939) who discussed the notion of *social forces*. Lewin's account of spatial influence has been used to model and simulate how pedestrians coordinate conflicts of space, like when passing a door opening and how they form lanes (Helbing and Molnár, 1995).

In the following I will introduce some relevant notions about spatiality and then turn to the analysis of spatial influence that has been carried out on the corpus data (see Chapter 5). The last section of this chapter is devoted to a discussion of how

[1] which can be described as a fire place for guests, found in some villages in northern Namibia.

spatially oriented actions can be used to actively influence the spatial behaviour of humans through spatial prompting, incorporated as a design element in natural language user interfaces for communicative service robots.

This chapter is primarily based on findings in corpus data and the goal is to conceptualise spatial prompting and to motivate the possibility of using it as a design element in human-robot communication. Building systems that has the capability to use spatial prompts is the next natural step, but falls out of the scope of this thesis.

8.1 Spatiality in human-robot interaction

In order to actively influence the spatial behaviour of humans, a robot is dependent on different types of knowledge. The robot needs to know the goals of the current activity and how they relate to the spatial dimension of the current situation. The spatial understanding of the current situation is partly dependent on knowledge about spatial phenomena, interpretations of the environment based on sensor information, and information given through communication between the robot and its users. Linking spatial representation and natural language has been addressed in different ways and provides a large and challenging area of research, especially the research that concerns the cognitive and linguistic dimension of space. In the following I will discuss concepts and notions from two main perspectives: spatial knowledge and social spatial behaviour.

Spatial knowledge

Research on spatial knowledge is focused on representation and reasoning from a cognitive perspective. The main application of this is to understand the way space is represented, understood and expressed using natural language. Spatial knowledge can be used to understand and generate linguistic descriptions based on spatial features in the environment and relations between objects. This has been investigated by Moratz et al (2008; 2003) in the field of Qualitative spatial reasoning, where the overall problem is to abstract metrical details of the physical world.

There are two basic directions in this research field: Topological reasoning about regions (Renz and Nebel, 1999) and reasoning based on orientation of points

defined in coordinate systems. These coordinate systems may be relative or absolute (Moratz and Ragni, 2008).

Another area of research that has implications for human-robot interaction concerns how robots learn representations of space. Spatial representation and learning for robotics has been approached by developing cognitive models of the environment, like the notion of the *cognitive map* as a Spatial Semantic Hierarchy (Kuipers, 2007). The hierarchy represents space on different levels of abstraction. The perspective, frame of reference and domain knowledge of humans can be used to augment this learning process (Topp et al., 2006).

The spatial semantic hierarchy describes common-sense knowledge of spatial concepts in terms of large-scale space. According to Kuipers (2007), the "large-scale" refers to space that conceptually is at a scale larger than the sensory horizon: the cognitive map. Small-scale space on the other hand is built up directly based on perceptual information and is usually referred to as a local perceptual map (Kuipers, 2007).

Spatial language

According to Levinson (1996) spatial communication can either use a frame of reference, some kind of coordinate system, or communication can be carried out without a specific frame of reference.

To communicate about space without assuming a frame of reference, Levinson (1996) describes the following ways: deixis (for instance, *"here", "there"*) , topological relations (spatial terms like *"on", "in-between", "beside"*) and named locations, using toponyms like *"X is at Y"*.

In cases where a frame of reference, some kind of coordinate system is used Levinson (1996) points out three strategies for referring to space:

- Intrinsic: using terms like *"in front of"* in relation to features of a background object.
- Relative: using terms like "left" or "right" and "in front of" or "behind" to refer to a position coordinates based on the speakers position related to the object.
- Absolute: using directions like "north" or "south", etc, to provide a fixed frame of reference.

For instance if a cat *"is in front of the car"* it is positioned between the speaker and the car in the relative reading. In an intrinsic reading of the example the cat is positioned in a location that is close to the front-side of the car, irrespective of the position of the speaker. If a cat is placed *"north of the car"* it does not matter where the speaker is located.

How a specific frame of reference is selected has been investigated by Tversky and Lee (1998) who suggest that proximity and salience of objects play an important role in which type of frame of reference is used. Other factors, which have to do with the domain in which the conversation is being carried out together with semantic and pragmatic factors (depending on the imminent goals and previous experience of the interlocutors) all contribute to the selection of a specific frame of reference. Natural borders, like the side of a room together with horizontal and vertical lines may also serve as reference frames in conversation. The human body, and its natural projected axes (head/feet, front/back, left/right etc) provide a frame of reference that is readily available to humans (Tversky and Lee, 1998).

Perspective taking, the ability to understand and communicate about space by taking the perspective of the communicative partner(s), is another type of phenomenon which has gained interest within robotics research. Trafton et al (2005) have investigated how cognitive models of perspective-taking can be integrated into an interface for a robot designed to collaborate with astronauts. The models allow the system to take the perspective of the human collaboration partner and are expressed as formal logical relations which allows traceability and integration with other approaches for symbol reasoning and planning. The simulations provide alternative representations of states of the world. This allows the system to take initiative to resolve ambiguity, when the system cannot fit one of its simulation models to the scene. Most of the utterances discussed by Trafton et al (2005) were referring to the functional relations of the object (for instance: *put the forward part of the spud into position*). Many utterances required some type of perspective taking and the speakers frequently shifted their perspective. Based on their findings Trafton et al (2005) argued that the robotic object representations, the reasoning and perception mechanisms should be as similar to those of humans as possible and integrated into a cognitive architecture. Furthermore it is necessary to apply heuristics and principles for collaborative activities that are similar to what is ordinarily employed by people, to match expectations.

Social spatial behaviour

There are several research approaches for human-human interaction regarding social spatial behaviour that are relevant for spatial management between humans and robots. For studies of social distances in human-robot interaction Hall's work on interpersonal distances (Hall, 1966) is relevant. Hall distinguished between four different distances: intimate (0-1.5 ft), personal (1.5-4 ft), social 4-12 ft, and public (> 12 ft). The social distances reported by Hall varied with the social activity and with the cultural background of the interlocutors.

Walters et al (2005a; 2005b; 2007) studied approach directions and distances in different conditions (the robot approaches a participant who is seated, positioned against a wall, or standing in the middle of a room). These approaches were intended to model scenarios where a hand over of an object takes place. Walters et al (2005a) found that approaches from either left or right were rated more positively than approaches from the front or from the rear. Approach distances have been studied by Koay et al (2007) who reports that for a task involving a behaviour for handing over an object the distance when the participants felt comfortable was about 60-70 cm[2], something which falls into the span defined by Hall's as *personal* distance.

Spatial formation

Another observation concerning the social use of space which is relevant for robotics was investigated by Kendon (1990) who introduced the concept of spatial formation. He made the observation that people dynamically grouped together in clusters, along lines and in circular shapes. He analysed this in terms of what he referred to as an F-Formation system.

Central to the notion of an F-Formation system is the shared space, the so called o-space, or the transactional space which is can be described as an area, normally located in front of the interlocutors, where the interaction is conducted. Clark (1996) refers to this space as the workspace. It is also in this area where perceptual co-presence can be established between speakers (Clark and Krych, 2004; Emery, 2000).

[2]66.8 cm (Koay et al., 2007)

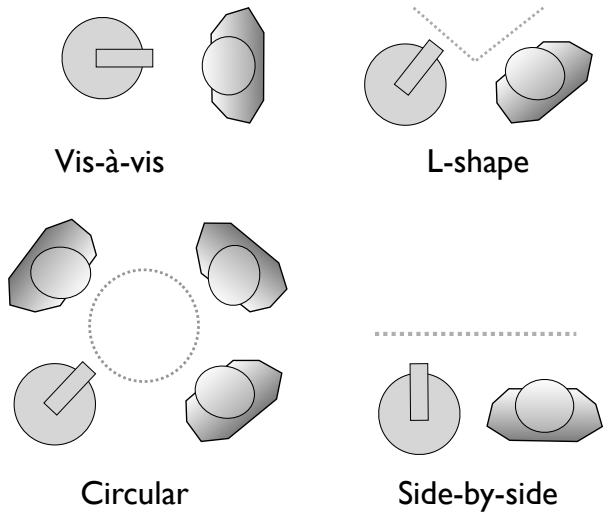


Figure 8.1 *Different possible F-formations involving a robot and a human based on Kendon's (1990) human-human counterparts.*

There are several ways in which interlocutors can relate to transactional space (and thereby to each other). Kendon (1990) argued that some patterns can be said to be prototypical for certain joint activities. In the following examples (see Figure 8.1) the F-formation patterns have been translated to a counterpart involving an embodied robot:

Vis-à-vis – robot and human facing each other.

L-Shape – robot and human on the legs of a virtual L-shape figure between them.

Side-by-Side – robot and human facing an outer edge together, standing in parallel beside each other.

Circular – several agents are surrounding a common area, forming a circle.

The formations that arise in the dynamic process of spatial formation are characterised in the F-Formation system in terms of their shape, for instance the L-shape which describes the relation when two participants have a common visual focus.

Another type of arrangement is the Vis-à-vis formation, where people are facing each other. The side-by-side arrangement is possible in a situation where peo-

ple are sharing access to a surface, like a white board. Circular arrangements involving several people engaged in conversation is also a possibility.

One example in robotics where transactional space is used as a fundamental configuration in the interaction is the tutoring system described by Sidner et al (2004) which was equipped with an interface robot shaped like a penguin. The robot penguin could turn its head and direct its gaze to objects located on a table in the environment. The dialogue system that managed the interactions with the user system used the robot penguin r both as a pointing device; for instance by looking at objects; and as a feedback device that gave conversational feedback, by looking at the user to influence turn-taking behaviour.

Another example that can be analysed in terms of F-formations is the embodied virtual characters[3]. With few exceptions they all share one common feature, they engage users in what we can call a *Face-to-Face F-formation*, using the terminology of Kendon (1990). This means that the system and the user share a transactional space which is located both in the virtual and the real world at the same time. One of the early examples of the idea of an intelligent environment with anthropomodal interaction modalities was described by Bolt (1980). Using Bolt's Put-that-there system as an inspiration, Thórisson (1997) created a similar environment with an embodied virtual character equipped with a mouth, moving eyebrows and gaze. These actuators provided conversational feedback (like back-channel feedback), attentional and deictic functions (using the eyes or the hand) and emotional displays (for instance smiling). This spatial arrangement has been used in several systems involving virtual characters, the Cloddy Hans system (Gustafson et al., 2004), REA (Cassell et al., 1999), Olga (Beskow et al., 1997) and the Pixie system (Gustafson and Sjölander, 2002).

In the REA system and the Gandalf system the users are shown objects slightly to the side (Cassell et al., 1999; Thorisson, 1997). In the case of the Cloddy Hans system and to some extent in the REA system, the space also extends to conversations around objects located in the room behind the character (Cassell et al., 1999; Gustafson et al., 2004). This configuration has similarities with the L-shape, as the character stands to the side and displays items located in what visually appears as

[3]For excellent overviews of this field please see (Bell, 2003; Gustafson, 2002)

a transactional space, extending from the human standing in front of the character to the virtual 3D environment on the side of the character.

The Valerie system (Gockley et al., 2005) is focused on long-term relations and the use of narratives that evolve over time. By detecting persons passing and standing in the proximity of the robot, the robot can direct its gaze towards people and say things to people that are in the vicinity of the system (Michalowski et al., 2006).

In the Valerie system a model of engagement was used to engage visitors. The model used four spatial regions to infer the engagement status of a person interacting with the system. These regions are schematically depicted in Figure 8.2:

- Present: people far from the robot or people passing fast.
- Attending: people who are coming closer to the robot.
- Engaged: people who are standing close to the robot.
- Interacting: people who are actively interacting with the robot.

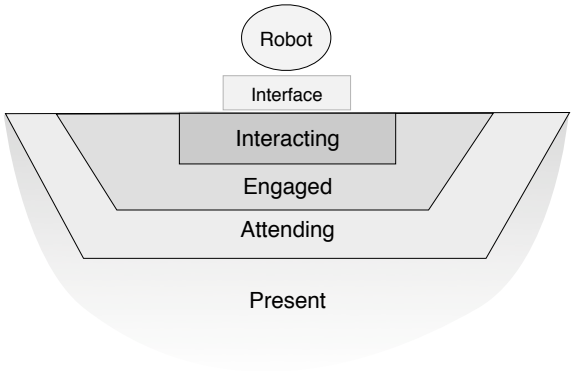


Figure 8.2 *A model of engagement space after (Gockley et al., 2005; Michalowski et al., 2006)*

Another way of using spatial distance to model communicative behaviour is the notion of a friendliness map proposed by Tasaki et al (2005). In their system, a humanoid robot in a fixed position (similar to Valerie system above) they varied the parameters for what type of communication the robot should engage in, based on a model of spatial proximity inspired by Hall's proxemics (Hall, 1966). The robot

responded to tactile communication, speech input and face detection in intimate space, speech input and face detection in personal space and face detection in social space. The robot also made decisions on what type of activity it would engage in, by attempting to play a game with users present in personal space, selecting the direction where the highest "friendliness" is found.

Communicative effects of body movements

In the corpus data described in Chapter 5 we observed that the robot's spatial actions involving the whole body had a communicative effect. In a related study on the same corpus Hüttenrauch et al (2006a) found that humans adjusted their spatial distance in relation to the imminent task, which was to show objects and locations to the robot. These phenomena were interpreted as a preparation for a new episode in the overall task. These results can be seen in the light of research on communicative functions of the body in human-human interaction.

Arguably human body movements have a communicative effect. One account is given by Gill et al (1999; 2000) who investigated the communicative effects that participants achieve by using nonverbal behaviour, focusing on the functional rather than the morphological perspective of nonverbal behaviour. One such function is termed *focus* which, according to Gill and Borchers, is a meta-discursive function that signals a shift in the center of attention in the discussion, for instance, a shift in body posture with the same meaning as the utterance *"I am going to focus on this spot"*. Gill and Borchers' view of the role of *engagement space* (Gill and Borchers, 2003), defined as the "the aggregate of the participants' body fields for engagement", also includes *pragmatic* dimensions of spatiality and is used to explain how spatial body moves can be understood as having communicative functions. An engagement field is based on the commitments by the participants to be bodily involved in the activity at hand. The body field is variable and depends on to what degree the participants feel comfortable or uncomfortable. Gill and Borchers (2003) provide an example where if "one person moves their hand over into the other's space, and that person withdraws their hand, this indicates that the *contact* between these persons is disturbed". This disagreement, or discrepancy in the body field makes it necessary to reconfigure the body until that a mutual feeling of "sharing an engagement space is re-established" (Gill and Borchers, 2003). This phenomenon has also been characterised by Schegloff (1998) using the notion of

Body Torque, indicating that a displacement or reconfiguration of the engagement space on one part projects a change in the conversation, leading to a new configuration.

The scenario that Gill analysed concerned interactions in front of a large electronic display (a SmartBoard). To explain spatial interaction in the scenario described by Gill (2003), proposed that activities of the interlocutors could be attributed to specific spatial zones:

- Reflection zone: one person is acting at the surface in focus and the other is standing back, observing the actions.
- Negotiation zone: the participants are "engaging about an idea and there is some movement or indication to access the surface, then this occurs in the negotiation zone".
- Action zone: where actions are performed. This involves direct physical contact with the surface.

Analysing body moves in human-human communication cannot be done in a straight-forward manner, as the joint creation of meaning is dependent on the situated and changing context. In the next section I will describe and discuss a set of observations regarding spatial influence in human-robot communication that may serve as a motivation for the possible use of spatial prompting as an active design element in human-robot communication.

8.2 Spatial influence in the corpus data

By analysing the video corpus from the Home Tour scenario we identified and described instances where the robot movements or verbal actions appeared to influence the actions of the user. The examples reflect three different ways in which the robot actively influences the user to act:

- Spoken and conventional ways of inciting spatial actions, such as action directives: "please stand in front of the camera".
- Positioning that leads to spatial formations, such as the L-shape.
- Spatial actions that trigger communicative behaviour, issuing a gesture (like a "stop" gesture) or releasing change projected by body torque (for instance, returning to an upright position when robot provides verbal conformation).

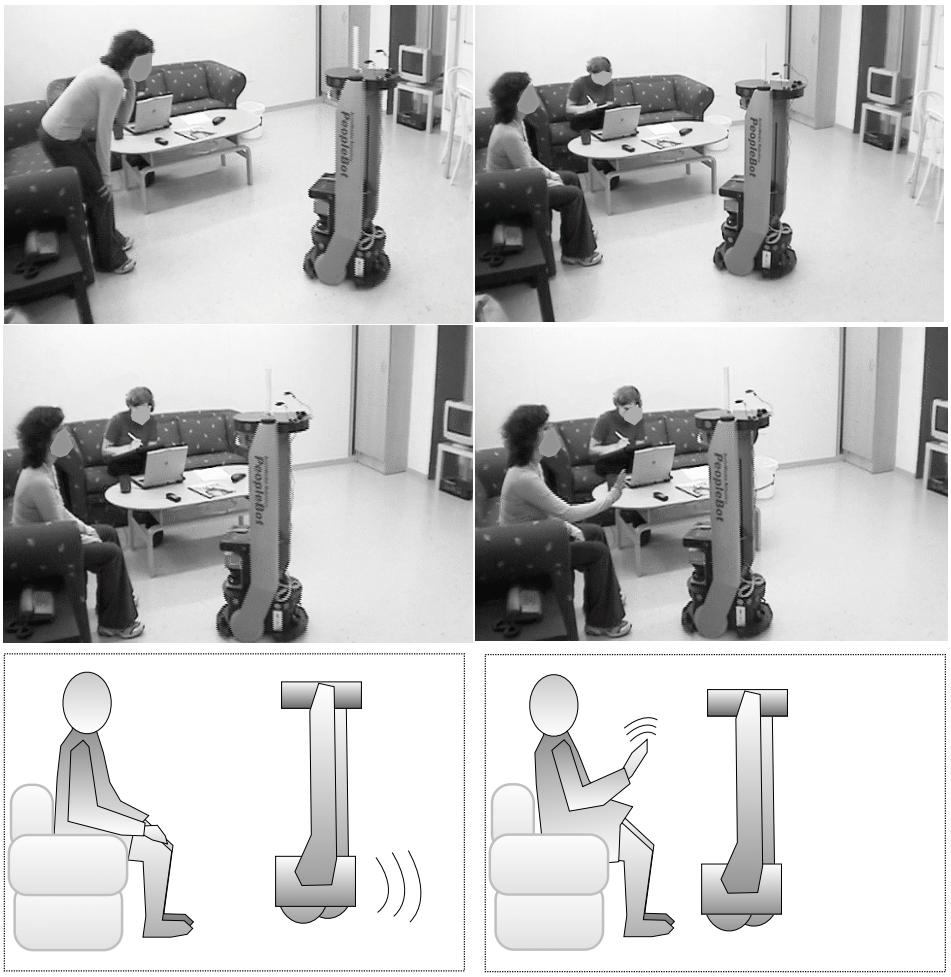


Figure 8.3 *Spatial action that influences communicative behaviour. The two upper images display how the participant sits down after noticing that the robot is starting to follow. The two lower rows of images shows how the robot approach triggers a "stop" gesture, which can be seen as an attempt of the user to control the situation, by raising an initiative. The lower two sketches visualises the same situation in a more schematic way.*

An example of how the (non-verbal) movements of the robot platform can trigger actions of the user is the following. When the user has commanded the robot to follow (by saying "Follow me"), the user sits down and waits as the robot is ap-

proaching. During the approach the user raises the arm and *displays a "Stop" gesture*. It appears as if the robot comes too close; perhaps crosses the border between a social to an intimate distance, in terms of Hall (1966) or triggers a behavioural reaction as the robot breaches a territorial border upheld by the user (Sack, 1986).

On the other hand we might interpret the raising of the hand in a "Stop" gesture as an indication to the robot that this is an advantageous position for the task at hand, which means that the "Stop" gesture is displayed as part of a joint communicative goal (using the terminology according to Clark 1996).

Figure 8.4 *An example of how communicative actions and spatial configurations of the robot are interrelated in different modalities.*

Another example where verbal productions and body movements interplay is depicted in Figure 8.4. In the moment that passes before the example the user has acknowledged that the robot has completed the task of finding an object (by establishing a common reference to the object located in front of the robot). Then the user bends the upper body forward and looks into the camera of the robot, while uttering the command: *"Now go to the telephone"*. This can be seen as a way of establishing an engagement space in-between the robot and the participant. Then, when the robot confirms the request by saying *"Going to telephone"*, the user changes into an upright position, something which can be interpreted as the bodily action of acknowledging the confirmation of the robot. This also ends the sequence of interaction and this in turn no longer makes the necessity of maintaining the engagement space between the participant and the robot.

Eye contact is maintained during the whole sequence. My analysis of this is that the user is attempting to acquire (visual) attention on the part of the robot. It can be argued that the moving camera of the robot provides a bio-mimetic display that makes the user assume a transactional space, using the terminology of Kendon (1990), located in front of the robot. The bow-forward posture can be attributed to an attempt to acquire contact by establishing a transactional space between the participant and the robot. The reason why the user is bowing forward is because she is constrained by the affordance provided bye the gaze direction of the camera. In order for the spatial configuration to define a transactional space, the user and the robot must be in a state of communicative contact. As the robot is non-responsive in this respect, the user adapts to the situation by bowing forward and seeking gaze.

When analysing the data from the Wizard-of-Oz study for the Home Tour in the Cogniron project (see Section 3.4) we found that there were some things regarding the interaction for getting the robot to follow the user that caused some concern. It appeared as the actions of the robot seem to trigger, or influence the behaviour of the user. Initially we believed this to be something of a flaw of the study, namely that the wizard used preemptive actions to cause the user to act. But when inspecting the data more closely we found that the robot influencing the behaviour of the user was something that was recurring throughout the different sessions. I have already introduced the example in Figure 8.3 where communicative behaviour (raising a hand to a form a stop-gesture) is triggered by the robot approaching the user. In the example depicted in Figure 8.5, the robot is doing something that was unexpected. Due to a glitch in the programming the robot rolled straight until it stopped by slamming into a wall, where it stopped because the bumper switches triggered the motors to shut off. The fact that this was an erroneous behaviour of the robot was not known to the user, and while the robot was moving in a straight line, the participant is trying to maintain a face-to-face configuration and moves closer up to the point that it appears as the participant is trying to wedge herself between the wall and the robot. I have interpreted this as an attempt to communicative with the robot rather than manually stop it, because the participant is leaning forward while looking (very closely) into the robot's camera. She does not attempt to physically grab the robot, keeping her hands on her knees.

The will to adapt to what appears to users as a communicative robot leads to assumption that we can influence the spatial behaviour of the users. It is well known in literature that communicative actions can be analysed in terms of discourse obligations (Traum and Allen, 1994), for instance in a conversation between A and B, a request X by A is followed by the obligation to reject X or accept X. It seems that the adaptation that we have seen in the examples can be viewed as a way on the part of the user to fulfil such obligations, but through spatial actions. As spatial adaptation was not the primary topic of study in the Wizard-of-Oz scenario the phenomena we observed seemed more like a side-effect of the robot and the participant trying to achieve a common goal. It is an interesting challenge to investigate how we can we can achieve *spatial* adaptation through communicative acts in a more deliberate way and use it in the design of practical system.

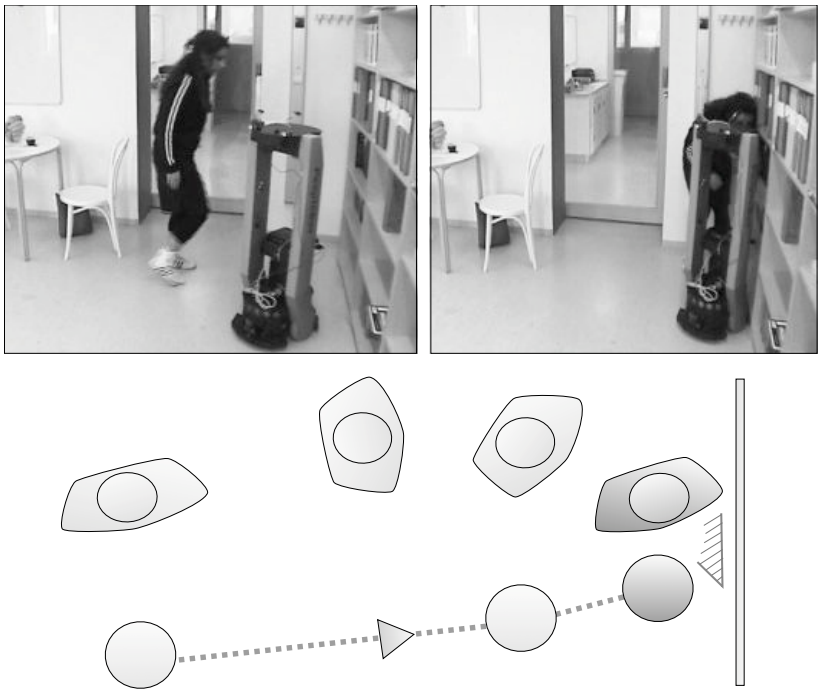


Figure 8.5 *The user attempting to maintain gaze and grab the attention of the robot by wedging in-between the robot and the wall.*

8.3 Spatial prompting

In this section I will turn to discuss how spatial influence can be used in design of human-robot communication, introducing the term *spatial prompting*.

Used as a verb the word *prompt* means *to move* or *incite to action*. I have chosen this tentative definition:

> *A spatial prompt is a communicative action that incites someone to spatial action, or a spatial action that incites someone to communicative action*[4].

These cues may be verbal, gestured, or performed using the whole body. They may also be auditory, visual or tactile displays, that provide directions for the user's spatial action. It is important to keep in mind that this is a practical definition that is to be used in the discussion of design of physically situated human-robot interaction. In terms of interaction design the physical appearance and the robot's set of behaviours provide the means of defining a *spatial prompting strategy* to be used by the robot.

Another approach is to explore strategies for spatial adaptation on the part of the robot rather than spatial prompting, for instance by taking conventional information into account as a heuristic for what can be seen as appropriate movement, like keeping to the right while passing humans (Pacchierotti et al., 2005). This is also well exemplified in by Zender et al (2007) who employ a spatial adaptation strategy for person following which uses Hall's notion about personal distance to trigger the follow behaviour (Hall, 1966). When the user leaves personal space (about 1.2 metres) the robot starts its following behaviour.

In terms of design we can do this by combining verbal and non-verbal communicative actions together with movements and positioning of the robot into multimodal *spatial prompting strategies*. A potential design that is modelled after the empiric finding, visualised in Figure 8.3, would be to trigger the use of stop gestures by *increasing* the speed of the robot slightly when moving closer (and then come to a sudden halt). It is likely that this robot behaviour would not be acceptable to users, but is illustrates the possibility of using empirical finds to inform design.

[4]This definition rules out cases where spatial action influences spatial actions (activities such as "dancing", "boxing", etc) or when communicative actions influence communicative actions (activities involving verbal exchanges only).

Cases of spatial prompting in service robotics

Even if the term spatial prompting is new, the practical use of robot behaviour to influence spatial action in human-robot interaction is not new. On the contrary, there are several examples in the literature, but it seems that the terminology and the overall framework for describing the phenomenon is largely missing. The examples of use of spatial prompting discussed in the following can be therefore be seen as ways of dealing with a specific interaction situation rather than attempts of implementing a principle based solution for dealing with managing spatial interaction. When revising the literature over the last decade in human-robot interaction some types of themes or activities for which spatial prompting seems important emerge:

- Crowd control, controlling behaviour of groups of people.
- Resolving issues related to robot following behaviours.

Crowd control

I appears as the problem of "crowd control" has evolved from navigating through crowds to more complex schemes for managing group behaviour to prepare for further interaction.

One example of use spatial prompting to facilitate navigation through crowds is the museum robot Minerva (Thrun et al., 1999). During it guided tours the robot asked people to step out of its way. This could be done friendly, when the robot is in an emotional state corresponding to being happy. It could also be done in less friendly manner by uttering "You are in my way" while simultaneously displaying a frowning facial expression, corresponding to the emotion "angry".

To control the behaviour of a group of people several approaches have been proposed. A system that uses a variant of spatial prompting to control a crowd with a team of mobile robots is described by Martinez-Garcia et al (2006). Their system is inspired by the behaviour of sheep dogs, because they noted that the behaviour of dogs, who engage in the herding of sheep involves a combination of behaviour such as barking and running towards the flock. The system described by Martinez-Garcia et al (2006) only uses the implicit signals given by the robot's motion and trajectory. The intended use for the system to use it in as a team of robots to a) guide, b) group or crowd, or c) intercept the group of people in focus.

Guiding means that the group of people follow the robot. Grouping entails the concentration of the group of people by influencing them. By intercepting users that try to leave their group, the robots strive to keep the group together. Martinez-Garcia et al (2006; 2005) only reports on work based on simulation studies (without user involvement).

Another perspective of controlling the positioning of people acting as crowds is provided by Shiomi et al (2007) who propose a set of verbal spatial prompts for "Group Attention Control". In the situation they describe, a humanoid robot[5] situated in a science museum, the verbal spatial prompts are used to influence the people crowding in front of the robot. In a Wizard-of-Oz study they performed, the robot detects different spatial configurations and reacts according to fixed schema of verbal prompts. When the robot detects that the visitors should be in front of the robot, it says "Please stand side by side in front of me". Likewise when the distance is not optimal the robot can prompt visitors to "Come closer", "Move back a little". Shiomi et al (2007) argues that this creates what is characterised as a "social situation", perhaps better understood in terms of spatial arrangements using Kendon's (1990) terminology.

Managing people following

People following is another activity for which spatial prompting seems to play an important role. Gockley et al (2007) describes how their robot, intended to follow people in a way that which is perceived as more "natural"[6]. I will not discuss the movement pattern here, instead turn to the way verbal prompts are used in the follow behaviour. In their system, upon detection of a user, they used an instruction, which in effect was a spatial prompt: "Start walking, and I will follow you". Since the task for the participant in the scenario was to lead the robot there were also prompts for making the participant continue walking: "Don't stop!" and "Why are you stopping?". The system also uttered more general comments to appear more encouraging and sociable, like "You're doing great!" and "Keep it up!". When the user was lost by the system, the robot uttered "I've lost you!". This should be seen in contrast to for instance the dialogue design used in the Wizard-of-Oz study described in Chapter 5 where the spatial prompt "Please get in front

[5]RoboVie (Kanda et al., 2002b)

[6]I understand this use of 'natural' as "more like a human would follow"

of the camera!" was used to reestablish the following. A careful interpretation is that "I've lost you!" can be understood as a indirect speech act requesting the user to get back in front of the robot in order for the robot to resume the following behaviour.

To handle situations where the follow behaviour cannot start because the user does not move away are common in systems that employ person following like the one sketched by Zender et al (2007). In Mahani (2006) this was characterised as a "deadlock situation" and a spatial prompting strategy was used to overcome the deadlock. The strategy involved the use a verbal spatial prompt ("you are too close to me, you have to move a bit further and then I can follow you") together with the display of a "wiggling" motion, where the whole robot body moves towards the user and then displays a wiggling motion from side to side until the user moves away far enough for the following behaviour to start. Mahani (2006) reported that these strategies were effective in resolving the deadlock.

Uses for spatial prompting in robot design

Given the examples of spatial influence that were found in the corpus and the robot designs listed above, where the intention is to spatially influence humans, it is now possible to summarise a set of general uses for spatial prompting:

> *Eliciting spatial actions of the user:* This can be spatial adaptation, for instance moving out of the way, keeping to the right. It may also be coordination of spatial actions, for instance maintaining a certain speed, stopping and starting movement. The robot may also position itself in such a way that it is possible for the user to position herself so that a spatial formation is achieved, like Vis-à-vis, L-shape or Side-by-side. By positioning themselves according to an F-formation, the robot and the user have defined a transactional space, something which can be taken into account when focusing the perceptual mechanisms of the robot.
>
> *Temporal alignment of communicative actions of the user:* This may be actions of the robot that trigger communicative behaviour of the user at a specific time. This could be observed in the example where the stop gesture was issued when the robot came too close (Figure 8.3). Another possible example of a spatial prompt that may influence the temporal alignment of

communicative actions is when the robot body starts or stops moving. This can be can seen as feedback and evidence of understanding given at a specific instance in time. The system could then expect a new command or action related to the last action.

Establishing of joint perceptual attention or referencing the environment: By turning the body the robot can mimic a display of communicative attention towards some specific object or location. This may, under some felicitous circumstances, be interpreted by the users as joint perceptual attention.

In my view, in order to use spatial prompting as a design element in human-robot communication, it is necessary to incorporate models of spatial relationships into dialogue management systems allowing for the display and interpretation of spatial prompts as an integrated part of multi-modal dialogue. The prospects of extending spatial prompting beyond crowd control and people following, and using a more principle based approach is being explored by for instance Kaindl et al (2008) who sketch a dialogue model for a system that treats spatial prompts on the same level as Speech Acts. They describe an example where in a scenario involving interaction with a shopping trolley the (verbal) acceptance of the request to go to a product is emphasized by the movement of the trolley: "[it] will slowly start to move, to emphasize its acceptance of going to this destination together with the user, at the same time as uttering the Accept communicative act".

8.4 Chapter summary

In this chapter I have introduced the concept of *spatial prompting* and motivated its use in the design of human robot communication. Spatial communication in human-robot interaction can be analysed using several different research frameworks. Approaches taken from cognitive psychology have been focused on the way humans and robot can understand spatial knowledge and reasoning about spatial phenomena. The cognitive approaches typically involve attempts to relate spatial knowledge to spatial language. The understanding of social spatial behaviour in human-human interaction has also influenced the research on human-robot communication.

Using the corpus data I have shown examples of how communicative actions of the robot influence the behaviour of the humans that interacted with the robot

in the home tour scenario (described in Chapter 5). Another thing that has been discussed in this chapter is that robot actions seem to trigger communicative behaviour of humans. I also found that robot communicative actions seem to trigger spatial behaviour and body movements of humans. These body movements, can be analysed in terms of proxemics, or other types of frameworks that attempt to account for the social use of space.

To summarise this I argue that it is possible to use spatial prompting strategies to influence the spatial and communicative behaviour, for instance to trigger users' spatial actions, adaptation to F-formations, temporal alignment of communicative actions and displaying attentional references to objects and locations.

CHAPTER 9

Concluding discussion

The research presented in this thesis focuses on interaction design and evaluation of human-robot communication. The interaction design in this work has used human-to-human communication as a source of inspiration and theoretical backdrop.

In the introduction research questions centred around three themes were posed. The first theme concerns design of human-robot communication asking the question: what would be an appropriate communication design for a service robot? The second theme in this thesis concerned the question of how to evaluate interactive systems for human-robot communication, and more specifically, how this can be done by analysing miscommunication using a corpus collected in realistic use scenarios. The third theme that has been addressed in this thesis has emanated from the work with corpus-based evaluation, and is concerned with the question of how phenomena of spatial influence can be motivated empirically and used as design elements for communicative service robots.

9.1 Evaluation of human-robot communication in realistic scenarios

The approach in this work has been focused on the design of human-robot communication using aspects and traits of human communication as a basis for development of robot prototypes. These prototypes have been intended to carry out physical service tasks for users in realistic scenarios, although the system functionality has either been simulated or implemented in real, but limited, working prototypes.

The overarching goal has been to study robots that employ natural language in an instrumental way to solve physical service tasks, rather than engaging in social conversation. To this end the robot prototypes that I have worked with have been "functionally designed" (Fong et al., 2003a) because they are designed to employ strategies of social communicative behaviour, but without attempting to model the mental and biological inner workings of a human.

To answer the research questions the communication design for two autonomous service robots has been investigated. The first instance of communication design concerned the development of a natural language user interface for the office robot Cero. The activity of users in the investigated scenarios can be described as specification of long-range and directive goals through the use of natural language utterances like "go to the kitchen" or "move forward". Directive goals could either be expressed verbally or through conventional gestures.

The second instance of communication design that was studied involved service discovery and configuration for what can be characterised as a "Cognitive companion", a robot with the capability of exploring a home or a workplace while interactively engaging humans in conversation to build a representation of its environment. The information provided verbally by the users in the scenario were names of objects and places that were contextually situated simultaneously through deictic gestures, a type of information that was intended to support future service tasks.

To answer the research question how we can analyse and evaluate the quality human-robot communication, it is useful to recapitulate the way evaluation has been approached. On the one hand we have developed multimodal interfaces according to what can be viewed as a product development cycle. On the other hand we have been involved in a research process that goes beyond interface development. The evaluation activities can be seen as two intertwined processes. The outcome of a product development process is a kind of product, a research prototype, whereas the outcome of a research process is documentation and data of user studies, e.g., in the form of analyses and corpus data, and importantly – increased understanding of human-robot interaction. Only when individual components are integrated in a robot prototype can they be evaluated with respect to the user experience.

The suitability of Wizard-of-Oz

Why do we enact interaction with an embodied robot using the Wizard-of-Oz technique rather than attempting to build a working robot from scratch? As this research has been carried out under the flag of the framework of user-centered design we are dependent on opinions and interactions of users. This made it necessary to create prototypes very early in the design process, something that was made possible with the use of the Hi-fi simulation using the Wizard-of-Oz technique.

Initially the anticipated practical outcome of the Cero project was a robot that could be controlled using a speech interface based on off-the-shelf components. Soon we learned that a way forward, when some technology is not yet sufficiently mature or are being developed in parallel, would be to use the Wizard-of-Oz technique to enable user studies involving persons that had little or no experience of interacting with robots.

This was also the reason for the choice of method when the communication design for the initial version of the dialogue system[1] for the robot used in the Cogniron project. To allow the study of human-robot interaction the Wizard-of-Oz technique, which was originally developed for the prototyping for uni-modal, screen-based natural language user interface (Dahlbäck et al., 1993; Kelley, 1984; Malhotra, 1975; Maulsby et al., 1993) has been extended to support interaction with situated embodied agents.

One of the initial and yet fundamental findings was the understanding that human-robot communication with service robots is a situated multimodal activity. This means that verbal and gestural utterances, body movements, spatial orientation and physical relationships of both the robot and the human need to be studied to understand human-robot communication.

Through my engagement in this evaluation process I have gained an increased understanding of what needs to be considered and what can be learned. The experience from carrying out studies using the Wizard-of-Oz technique has strengthened the following methodological perspectives:

- Gaining first person experience from being responsible for the enactment of a multimodal interface gives important insights into human-robot communi-

[1] This dialogue system was later implemented on the robot BIRON at University of Bielefeld (Li, 2007)

cation, something which adheres the observations made by Maulsby (1993). To this should be added that design activities focused on the practical enactment of the system by necessity becomes focused on solving practical problems related to interaction. First, the immediate consequence of this is that the focus is on dialogue design of systems that are *complete* in the sense that the wizard operators have to cope with different types of user behaviour when enacting the system. Second it has the consequence that the designer considers users that are real, in the sense that they in a short time will appear in person to interact with the robot. Even if we do not know these particular users in person, knowing that they eventually will meet the robot, provides a kind of social pressure on the designers similar to that of an actor performing on the stage.

- The second methodological assumption that now appears to stand on more firm ground, because of the use of Hi-fi simulation, is that studying interaction with realistic prototypes is necessary for human-robot communication. Partly this has to do with the fact that building working prototypes requires a large effort. To assess a system on an early concept stage it is necessary to at least provide an approximation of what to expect. It has been noted by Bannon (1991) that users need to be in a situation of "future use" to be able to provide comments about the system that is under development and thus not yet realised as a working prototype. In our case, as we have used the Wizard-of-Oz method, we have not only put users into the position that they are interacting with what appears as a future system, they have also been doing this under the assumption that the system is really working.

It is important to stress that from the perspective of designers a simulated system may be realistic but it is not real. The use scenario affects the way simulation studies can be set up and carried out. In the initial simulation study we carried out in order to assess the Cogniron Home tour scenario the interaction was confined to a single room. In the later studies of the Home tour scenario (cf., Topp et al. 2006) involved interaction stretching over the rooms and corridors of an whole office floor. In such a case, when the mobility of both users and systems becomes a topic for investigation, the complexity of setting up the scenario increases, something which limits the amount and manner in which use data can be collected.

Wizard-of-Oz allows for creation of a realistic corpus

In the introduction I also asked the question of how corpora of human-robot communication can be created and used in the design process. A prototype created to be evaluated in a simulation study is represented both as a mental script based on the imagination and knowledge of the designer and as a robot embodiment.

The enacted prototypes based on the Wizard-of-Oz technique offer a way of creating a situation that can be recorded on video. This allows the subsequent analysis of interaction. During the work with the corpus some general observations were made. First of all the heterogeneous character of the data means that what is to be recorded needs to be planned carefully. Unexpected relationships may be possible, or impossible to study depending on what has been recorded. For instance spatial behaviour and use of deictic gestures can be studied as a unit if synchronised video-recordings and laser range finder data are available in a corpus.

To be able to trust data collected in a simulation study the behaviour of the simulated components needs to be considered carefully with respect to the following dimensions:

- Degree of system realism: Are we going to simulate a realistic system? We must decide to what extent we are going to simulate system behaviours so that they appear as realistic to the user, e.g., by introducing system misunderstanding that appears to come from miscrecognitions.
- Degree of exploratory freedom: Are we going to simulate according to an algorithm? Decisions that are made along this dimension concern whether or not we should allow a completely free-form interaction style, or if we should restrict the task. This is what Maulsby et al (1993) refer to as being "true to the algorithm".
- Behavioural mimicry: how natural-like should the system be in terms of appearance and interactive behaviours and to what extent should the system imitate nature in terms of appearance? This may be represented as a scale ranging from an appearance which is similar to humans (e.g., Ishiguro and Minato 2005) with a conversational style of interaction that closely mimic human capability, to non-anthropomorphic appearance that uses a command-based style of conversation.

Clearly the Wizard-of-Oz framework is suited for simulating a full-fledged interactive service robot. When carefully designed, simulation studies will provide data about different aspects of human-robot interaction that would otherwise be unattainable until large efforts had been spent on the creation of a working prototype.

9.2 Miscommunication: observations and design implications

In this work communicative quality has been analysed in mainly two ways. First of all by designing, reflecting over and enacting a use situation with human-robot communication has provided hours of first hand experience from observing humans and robots interacting in a situated environment. The tacit knowledge, or what Maulsby et al (1993) would describe as "prior implementation experience"[2], that this has provided as been an important driving force for the hypotheses that have been developed in this work. Secondly, corpus data collected during these interaction sessions has been annotated and analysed with respect to the communicative behaviour in general, specifically focused on different types of miscommunication and consequences for the spatial behaviour of robot.

The user studies performed in the scope of this thesis have two things in common, first of all, every user role-played, in the sense that they were not using the robot to fulfil their individual high-prioritised needs. Secondly, the situations in which the interactions were carried out were created using the Wizard-of-Oz technique. The gain from this approach was that we could enact a scenario of human robot communication that was manageable from the perspective of the available resources and that still appears very similar to "the real thing", i.e., human-robot interaction with a working prototype.

How users engage in communication with robots

An important finding, that was initially hinted to us during the data collection phase, but subsequently became evident when analysing the material was that the common denominator for almost each and every person that participated in the studies was that they seemed to have two concurrent goals:

[2]Maulsby (1993) discusses that in order to setup and carry out a Wizard-of-Oz study, it is necessary to have prior experience from an earlier study. I also believe this to be the case. The first wizard study with the Cero system provided a good training ground for subsequent studies in this work.

- Solving the scenario work task.
- Explore the robot's capabilities.

Communicative- and other actions that are used to control, adapt to and influence the behaviour of the robot needs to seen in the light of these overall goals. Through the live sessions and the corpus analyses I have made the following overall observations regarding the communicative behaviour of users:

- Users' initiation, continuation and upholding of the communicative activity is used to drive the interaction towards the main task goal.
- Users overtake service actions that the robot is incapable of performing.
- Users explore the robot capabilities once they understand that the robot can perform physical actions.
- Users continuously monitor the behaviour of the robot.
- Some phrases are identified and used as safe commands – a command that provides an easy way out for or the user when miscommunication makes interaction too difficult to cope with. The consequence of this is sometimes that the user leaves the task that was originally intended.

Feedback and grounding plays an important part in understanding how the user perceives the robot. A large portion of miscommunication that caused frustration was related to the lack of contact feedback in the communication.

The role of feedback

The preliminary investigations into human-human communication in Chapter 2, which were used to design prototypes to test with users (Chapter 4) made it clear that communicative feedback plays a central role in the design of human-robot communication. Subsequently by experiencing the interaction, through the practical tests carried out with the Cero robot and the enactment and analysis of the Wizard-of-Oz studies described in previous chapters I have realised that the capability of providing appropriate feedback is a crucial component of a communicative robot.

In the work with the Cero system a feedback model that focused on the task-related aspects on communication was sketched. The feedback given in this design mainly concerned to which extent a task goal could be considered to be grounded.

This type of feedback is needed to ground the goals to be carried out in the system. To this end a dialogue system for a service robot is no different from a dialogue system for an information-based task, like flight booking (Larsson, 2002) or route planning (Allen, 1995) where task-related feedback plays a major role.

In the scenarios investigated in this work there is another type of feedback that appears to have a more fundamental impact on the quality of human-robot communication, namely feedback related to perceptual processes and the willingness to interact. Continuous perceptual monitoring of several modalities is an integral part of the communication process between humans. The observations in the corpus indicated that users continuously monitor the behaviour of the robot and actively seek the perceptual attention of the robot. Humans have the biological capacity of monitoring their partners when engaging in interaction, whereas robots require specific system components and a model of perceptual status, both of itself and of the user. Information that can be derived from sensor information available in the system, like sketched in Chapter 7, can be used to give feedback on the perceptual and attentional state of the system. This feedback ranges from conventional feedback (displaying verbal and gestured utterances), signals that are intended to draw the attention, to physical actions of the robot, that signal perceptual activity and communicative availability.

9.3 Spatial prompting as a design element

In the introduction the research question whether robots could be designed to influence the spatial behaviour of users was posed rather tentatively. Previous research has indicated that the physical behaviour of a robot is tightly coupled with its communicative behaviour. In the scenarios we have investigated, the way the functions and behaviours of the robot have been interfaced is through communication. Hüttenrauch (2007) proposed that "the robot is the interface", in my view this should be understood as the communicative functions of robot's physical embodiment and actions should be seen an integral part of the repertoire of the communication capabilities of the robot. One argument for this is how actions of the robot can be seen as an evidence of understanding. The conceptualisation of the notion of *spatial prompting*, as an observable phenomenon and as a possible design element, has further strengthened this perspective on human-robot interaction.

The analysis of the corpus data showed that the users' spatial behaviour seemed to be influenced by the robot's behaviour, appearance or position. The circumstance that the physical position of the robot is taken into consideration by the users is perhaps uncontroversial – the interesting observation in the corpus material concerns the way the robot actively influences the behaviour of the users. The way the robot may influence the users' spatial behaviour range from communicative actions that are interpreted as conventional ways of inciting spatial actions, for instance through verbal prompts and gestures; to spatial positioning, that encourages spatial formations and spatial actions. Together these phenomena form a large possible design space that need to be considered when designing spatial communication.

Future work: a spatial influence theory

The notion of spatial prompting was something that emerged as a result of the studies carried out in the later stage in the research process leading forward to this thesis. This means that there are still several challenges that could not be addressed within the scope of this thesis.

Current dialogue systems that have been used for human-robot interaction typically do not include components for influencing spatial relationships through spatial prompting. There are indeed systems that include spatial models, but these typically concern representations of space that allow for interpretation or verbalisation of spatial representations, such as perspective taking (Trafton et al., 2005) or other spatial relations (Moratz et al., 2003; Skubic et al., 2004; Tellex and Roy, 2006; Tenbrink et al., 2002). To extend such approaches to handle spatial prompting we need to interpret spatial situations and understand the communicative effects of actions performed by the system, that influence the behaviour of humans.

My understanding of spatial prompts is that they can be modeled using approaches that are used in pragmatics, like conversational acts with similar status as Body moves Gill et al (1999; 2000) or Body torque (Schegloff, 1998). It has also been proposed that communication through movement can be modelled using classical Speech Acts as the basic unit of communication (Kaindl et al., 2008). We should note that spatial influence includes more than communicative actions. Also social behaviour, such as interpersonal distance (Hall, 1966) needs to be taken into consideration because active adaptations in social spatial distance may have similar

effects as a spatial prompt, e.g., by making the user adjust her position with respect to something that is experienced as comfortable.

To be able to interpret the communicative and spatial effects of robot actions we need to model spatial and communicative influence. Apart from the type of spatial knowledge that are already needed for handling spatial reasoning and verbalisation of spatial relations, a model needs to represent relationships such as:

- Social conventions for spatial positioning and adaptation.
- Communicative functions of spatial actions.
- Accounts of human behaviour that can be used to infer whether a spatially oriented action has been achieved successfully.

9.4 Communication design for service robots

In the introduction two questions regarding design of human-robot communication were posed. The first and perhaps general question concerns what we understand to be an *appropriate communication* design. The second, and more specific question concerns how we can *prevent miscommunication* through design.

In the previous chapters I have described how communication design has been approached for the two service robots that I have worked on. It would be presumptuous to assume that this work could provide answers to these questions without empirically evaluating a large range of different designs. Given the choice to use a simulation approach and analyse examples of use in detail I have formed an initial understanding of what a working communication design for service robots ought to comprise.

In the user sessions collected using the Wizard-of-Oz technique the users seemed to cope very well with the robot when it carried out the things that it was supposed to do according to the information that was given to the participants during the experiment sessions. Even if the users tried to do things that were clearly outside the scope of the task descriptions, such as showing objects located on the walls, or pointing out themselves (in the same manner as they would do with an object) and the robot provided the feedback “Cannot do that”, they seemed to accept this limitation of the system. In my view this was due to timely and relevant feedback. Likewise, when the robot did not respond or react to input, the participants displayed clear signs of frustration. The subsequent analysis of the type of utter-

ances that were used to address the system indicated that the linguistic complexity of the phrases could be handled by a natural language understanding component, if they could be detected by a speech recogniser. The technical development of robot perception has not been the focus as such in this work, but when it comes to design of service robots, the complexity speech and gesture recognition cannot be underestimated.

Evidently a communication design for a service robot needs to take the services that we intend to provide to users into consideration. These services decide the specific actions, domain knowledge and plans that the robot needs to handle from a task perspective. But to aggregate these parts of a system we need to establish:

- Perceptual capability that enable the robot to provide immediate feedback or information on the communicative status of the robot. This feedback (information) should give the users the sense of being aware of the robots ability and willingness to continue interaction.

To evaluate this capability I believe that the concept of approachability is useful to consider. Mehrabian (1967) introduced the concept of immediacy describe physical or psychological closeness during interpersonal communication. The positive qualities that we should seek in a communication design with respect to approachability is a system that users experience as *attentive, alert and available* for being influenced by communicative actions of its users. A system that is quite the opposite to what we should strive for would be characterised as *slow and non-responsive*.

I would like to stress that a robot that comprise the positive qualities listed above does not necessary need to be characterised as a friendly, social, or sociable robot. These concepts are indeed relevant, but like a human, a robot that is attentive, alert and available for influence does not need to be your friend or act according to moral or social conventions.

Let us turn to the question regarding prevention of miscommunication. There are many approaches for handling miscommunication, for instance disambiguation of references, miscrecognition or misconceptions, *once it has been detected* (c.f. Bohus and Rudnicky 2005; Skantze 2007). In the work with the Cero system and the subsequent analysis of miscommunication I have already hinted that perception capability was the main problem. What cannot be detected cannot be acted upon. No matter what strategies are used to handle miscommunication, such as

context-aware help, adaptive behaviour in the dialogue system (c.f. Li 2007), they are useless if the system cannot even detect that it has been addressed or that a user is present.

My answer to the question of how to prevent miscommunication is similar to what I suggest would be an appropriate design. First of all we should note that miscommunication provides users with boundaries to system capability. By focusing design efforts on a system that is approachable we implicitly design for what can be characterised as graceful failure. When we design for graceful failure, the goal is to avoid severe miscommunication, leading to breakdowns. Miscommunication cannot be reduced entirely. We should not shun the fact that miscommunication occurs, or as Martinovski and Traum (2003) put it:

> *[W]e don't need to work for a fusion between humans and machines by frenetically trying to eliminate any possible misunderstanding first, because misunderstanding is part of communication, no matter who the interlocutors are*

In my view the way to approach this is to reduce miscommunication, by designing the robot so that it remains approachable. If the robot can be approached by reacting on contributions from users at all stages during interaction many of the cases that causes miscommunication can be avoided. If the miscommunication can be treated as *misunderstandings* rather than *misperceptions*, methods that have been developed and used in dialogue systems should then be possible to apply to robot systems.

Future work: supporting approachability

A system that can understand and manage its attentional display behaviour will appear to the user as being attentive and responsive, something that is essential for the general usability of the system. With respect to communication design, two strands of research should be investigated:

How a contact and perceptual model can be designed and implemented in a situated system. The model should keep track of the of the user's attention, e.g., based on eye-gaze, spatial positioning, verbal and gestural behaviour. This means that we need to investigate how perceptual information from the robot's sensor components can be used as cues for interactive behaviour. Fusing these heterogeneous information sources is necessary to keep information about perceptual status updated during interaction.

The role of the concept of approachability should be investigated in a robotics scenario. How can we formulate usability criteria for natural language user interfaces for service robots to enable the construction of user interfaces that users perceive as being approachable, attentive and responsive during dialogue? How can we evaluate the effectiveness of a system that comprises behaviours for attentional display, i.e., how can we assess the approachability, responsiveness and attentiveness of the system?

9.5 Final thoughts

Many of the proposed designs of human-robot communication that have been brought forward in this thesis have not been taken up as requirements when it comes to design robots for practical use. This whole thesis can be seen as a way of putting focus on what I consider to be the real issues of creating high quality human-robot communication. Apart from the generic challenge for any service robot, namely to provide useful services to humans, the challenge of designing high quality communication design poses problems that need to be addressed by the robotics community. In my view this challenge is to design robots that are approachable, that provide relevant and timely feedback and provide the users with a sense of being in contact with the system at all time, meaning that the users is never left in a state where there are no options to proceed in the communication. The key to being a good service robot is to *do useful things – while staying in contact*.

Bibliography

Aberdeen, J., Doran, C., Damianos, L., Bayer, S., and Hirschman, L. (2001). Finding errors automatically in semantically tagged dialogues. In *Proceedings of the First International Conference on Human Language Technology Research*, pages 124–128.

Aberdeen, J. and Ferro, L. (2003). Dialogue patterns and misunderstandings. In *Proceedings of Error Handling in Spoken Dialogue Systems*, pages 17–21, Château d'Oex, Vaud, Switzerland.

Ahrenberg, L., Jönsson, A., and Dahlbäck, N. (1990). Discourse Representation and Discourse Management for Natural Language Interfaces. In *Proceedings of the Second Nordic Conference on Text Comprehension in Man and Machine*, Täby, Sweden.

Allen, J. and Core, M. (1997). Draft of DAMSL: Dialog Act Markup in Several Layers. webpage. http://www.cs.rochester.edu/ research/cisd/ resources/ damsl/ RevisedManual/.

Allen, J. F. (1995). The TRAINS Project: A case study in building a conversational planning agent. *Journal of Experimental and Theoretical AI (JETAI)*, 7:7–48.

Allwood, J. (1995). An Activity Based Approach to Pragmatics. Technical Report Gothenburg Papers in Theoretical Linguistics 76, Department of Linguistics, Göteborg University.

Allwood, J. (2002). Bodily Communication – Dimensions of Expression and Content. In *Multimodality in Language and Speech Systems*. Kluwer Academic Publishers, Dordrecth, The Netherlands.

Allwood, J. and Haglund, B. (1992). Communicative Activity Analysis of a Wizard of Oz Experiment. Technical report, Department of Linguistics, Göteborg University.

Allwood, J., Nivre, J., and Ahlsén, E. (1991). On the Semantics and Pragmatics of Linguistic Feedback. Technical Report Gothenburg Papers in Theoretical Linguistics 64, Göteborg University.

Alter, S. (2001). Which life cycle – Work system, information system, or software? *Communications of the AIS*, 7(17):1–52.

Andersson, M., Orebäck, A., Lindström, M., and Christensen, H. I. (1999). ISR: An Intelligent Service Robot. In *Sensor Based Intelligent Robots*, pages 287–310.

Asimov, I. (1968). *I, Robot*. Voyager.

Asoh, H., Vlassis, N., Motomura, Y., Asano, F., Hara, I., Hayamizu, S., Ito, K., Kurita, T., Matsui, T., Bunschoten, R., and Kröse, B. (2001). Jijo-2: An Office Robot that Communicates and Learns. *IEEE Intelligent Systems*, 16(5):46–55.

Austin, J. L. (1962). *How to Do Things with Words*. Clarendon Press, Oxford.

Bannon, L. J. (1991). From Human Factors to Human Actors The Role of Psychology and Human-Computer Interaction Studies in Systems Design. In Greenbaum, J. and Kyng, M., editors, *Design at Work.: Cooperative Design of Computer Systems*, pages 25–44. Lawrence Erlbaum Associates, Hillsdale.

Bell, L. (2003). *Linguistic Adaptations in Spoken Human-Computer Dialogues: Empirical studies of User Behavior*. PhD Thesis, KTH Royal Institute of Technology. TRITA-TMH 2003:11.

Bernsen, N. O., Dybkjær, H., and Dybkjær, L. (1998). *Designing Interactive Ipeech Systems: From First Ideas to User Testing*. Springer, London.

Beskow, J., Elenius, K., and McGlashan, S. (1997). OLGA - A Dialogue System with an Animated Talking Agent. In Kokkinakis, G., Fakotakis, N., and Dermatas, E., editors, *Proceedings of Eurospeech '97, 5th European Conference on Speech Communication and Technology*, pages 1651–1654, Rhodes, Greece.

Bødker, S. and Grønbäck, K. (1994). Cooperative Prototyping: Users and Designers in Mutual Activity. *International Journal of Man-Machine Studies*, 34:453–478.

Böhme, H.-J., Wilhelm, T., Key, J., Schauer, C., Schröter, Christofter, C., Gross, H.-M., and Hempel, T. (2003). An Approach to Multi-Modal Human-Machine Interaction for Intelligent Service Robots. *Robotics and Autonomous Systems*, 44(1):83–96.

Bohus, D. and Rudnicky, A. I. (2005). Error Handling in the RavenClaw Dialog Management Framework. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, pages 225–232, Vancouver, CA.

Bolt, R. A. (1980). Put-that-There: Voice and Gesture at the Graphics interface. *Computer Graphics*, 14:262–270.

Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G., and Berlin, M. (2005). Effects of Nonverbal Communication on Efficiency and Robustness in Human-Robot Teamwork. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Edmonton, Alberta, Canada.

Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of sound*. MIT Press, Cambridge, MA, USA. paperback.

Brennan, S. E. (2001). The Vocabulary Problem in Spoken Language Systems. In Luperfoy, S., editor, *Automated Spoken Dialog Systems*. MIT Press, Cambridge, MA, USA.

Brennan, S. E. and Hulteen, E. (1995). Interaction and Feedback in a Spoken Language System: A Theoretical Framework. *Knowledge-Based Systems*, 8:143 – 151.

Bugmann, G., Klein, E., Lauria, S., and Kyriacou, T. (2004). Corpus-Based Robotics: A Route Instruction Example. In *Proceedings of IAS-8*, pages 96–103, Amsterdam, NL.

Bugmann, G., Lauria, S., Kyriacou, T., Klein, E., Bos, J., and Coventry, K. (2001). Using Verbal Instruction for Route Learning. In *Proceedings of 3rd British Conference on Autonomous Mobile Robots and Autonomous Systems: Towards Intelligent Mobile Robots (TIMR'2001)*, Manchester.

Bunt, H. (1994). Context and Dialog Control. *Think Quarterly*, 3(1).

Bunt, H. (2000). Dialogue Pragmatics and Context Specification. In Bunt, H. and Black, W., editors, *Abduction, Belief and Context in Dialogue. Studies in Computational Pragmatics*, volume 1 of *Natural Language Processing*, pages 81–150. John Benjamins, Amsterdam.

Bunt, H. C. (1999). Dynamic Interpretation and Dialogue Theory. In Neel, M. T., Bouwhuis, D., and F., editors, *The Structure of Multimodal Dialogue*, volume 2. John Benjamins, Amsterdam, NL.

Buxton, B. (2007). *Sketching User Experiences – Getting the Design Right and the Right Design*. Microsoft Research, Redmond, Washington / Toronto, Canada.

Carletta, J., Isard, A., Isard, S., Kowtko, J. C., Doherty-Sneddon, G., and Anderson, A. H. (1997). The Reliability of a Dialogue Structure Coding Scheme. *Computational Linguistics*, 23(1):13–31.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjalmsson, H., and Yan, H. (1999). Embodiment in Conversational Interfaces: Rea. In *Proceedings of CHI'99*, pages 520 – 527. ACM Press.

Cassell, J., Bickmore, T., Vilhjalmsson, H., and Yan, H. (2001). More Than Just a Pretty Face: Conversational Protocols and the Affordances of Embodiment. *Knowledge-based Systems*, 14(1-2):55–64.

Cherry, C. E. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of Acoustic Society of America*, 25:975–979.

Clark, H. H. (1996). *Using Language*. Cambridge University Press, Cambridge.

Clark, H. H. and Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*.

Clark, H. H. and Marshall, C. R. (1981). Definite reference and mutual knowledge. In Sag, A. K. J., Webber, B. L., and Ivan, editors, *Elements of Discourse Understanding*, pages 10–62. Cambridge University Press, Cambridge.

Clark, H. H. and Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, 13:259–294.

Clarke, R. (1994). Asimov's laws of robotics: Implications for information technology. *Computer*, 27(1):57–66.

Cogniron (2003). COGNIRON, Annex 1 - Description of Work. EU Sixth Framework Program FP6-IST-002020, http://www.cogniron.org.

Dahlbäck, N., Jönsson, A., and Ahrenberg, L. (1993). Wizard of Oz studies - why and how. *Knowledge-Based Systems*, 6(4):258–256.

Dybkjaer, L. and Bernsen, N. O. (2000). Optimising the Usability of Spoken Language Dialogue Systems. *Natural Language Engineering*, 6(Parts 3/4):243–272. in J. v. Kuppevelt, U Heid, U. and H. Kamp (Eds.):Special Issue on Best Practice in Spoken Language Dialogue Systems Engineering.

Dybkjær, L., Bernsen, N. O., and Dybkjær, H. (1997). Designing Co-operativity in Spoken Human-Machine Dialogues. In Varghese, K. and Pfleger, S., editors, *Human Comfort and Security of Information Systems. Advanced Interfaces for the Information Society*, pages 104–124. Springer Verlag.

Efron, D. (1972). *Gesture, Race and Culture*. Approaches to semiotics; 9. The Hague: Mouton. Also published in 1941.

Ekman, P. and Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage and coding. *Semiotica*, 1(1):49 – 98.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, 24(6):581–604.

Ericsson, K. A. and Simon, H. A. (1980). Verbal Reports as Data. *Psychological Review*, 87(3):215–251.

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003a). A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3-4):143–166.

Fong, T., Thorpe, C., and Baur, C. (2001). Collaboration, Dialogue, and Human-Robot Interaction. In *Proceedings of the 10th International Symposium of Robotics Research*, Lorne, Victoria, Australia. Springer-Verlag, London.

Fong, T., Thorpe, C., and Baur, C. (2003b). Robot, asker of questions. *Robotics and Autonomous Systems*, 42(3-4):235–243.

Forlizzi, J. (2007). How robotic products become social products: an ethnographic study of cleaning in the home. In *HRI '07: Proceeding of the ACM/IEEE International conference on Human-robot interaction*, pages 129–136, New York, NY, USA. ACM Press.

Forlizzi, J. and DiSalvo, C. (2006). Service robots in the domestic environment: a study of the roomba vacuum in the home. In *HRI '06: Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 258–265, New York, NY, USA. ACM.

Fritsch, J., Kleinehagenbrock, M., Lang, S., Plötz, T., Fink, G. A., and Sagerer, G. (2003). Multi-modal anchoring for human-robot-interaction. *Robotics and Autonomous Systems, Special issue on Anchoring Symbols to Sensor Data in Single and Multiple Robot Systems*, 43(2–3):133–147.

Fussell, S. R., Kiesler, S., Setlock, L. D., and Yew, V. (2008). How people anthropomorphize robots. In *HRI '08: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 145–152, New York, NY, USA. ACM.

Gamm, S. and Haeb-Umback, R. (1995). User interface design of voice controlled consumer electronics. *Philips Journal of Research*, 49(4).

Gaver, W. W. (1991). Technology Affordances. In *Human factors in computing systems conference proceedings on Reaching through technology*, pages 79–84, New Orleans, LA USA.

Gibson, J. and Pick, A. D. (1963). Perception of another person's looking behavior. *American Journal of Psychology*, 76(3):386–394.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston.

Gill, S. P. and Borchers, J. (2003). Knowledge in co-action: social intelligence in collaborative design activity. *AI & Society*, 17(3):322–339.

Gill, S. P., Kawamori, M., Katagiri, Y., and Shimojima, A. (1999). Pragmatics of Body Moves. In *Proceedings of the Third International Cognitive Technology Conference (CT'99) Networked Minds*, San Francisco, USA.

Gill, S. P., Kawamori, M., Katagiri, Y., and Shimojima, A. (2000). Role of Body Moves in Dialogue. *International Journal of Language and Communciation, RASK*, 12.

Gockley, R., Bruce, A., Forlizzi, J., Michalowski, M., Mundell, A., Rosenthal, S., Sellner, B. P., Simmons, R., Snipes, K., Schultz, A., and Wang, J. (2005). Designing Robots for Long-Term Social Interaction. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems 2005 (IROS2005)*, pages 1338 – 1343.

Gockley, R., Forlizzi, J., and Simmons, R. (2007). Natural person-following behavior for social robots. In *HRI '07: Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 17–24, New York, NY, USA. ACM.

Goodwin, C. (1981). *Conversational Organization: Interaction Between Speakers and Hearers.* Academic Press, New York, USA.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32:1489–1522.

Green, A. (2001). C-Roids: Life-like Characters for Situated Natural Language User Interfaces. In *Proceedings of the 10th IEEE International Workshop on Robot and Human Interactive Communication Ro-Man 2001*, Bordeaux/Paris, France.

Green, A. and Hüttenrauch, H. (2006). Making a Case for Spatial Prompting in Human-Robot Communication. In *Multimodal Corpora: From Multimodal Behaviour theories to usable models, workshop at the Fifth international conference on Language Resources and Evaluation, LREC2006*, Genova, Italy.

Green, A., Hüttenrauch, H., Norman, M., Oestreicher, L., and Severinson Eklundh, K. (2000). User-Centered Design for Intelligent Service Robots. In *Proceedings of 9th IEEE International Workshop on Robot and Human Interactive Communication Ro-Man 2000*, Osaka, Japan.

Green, A., Hüttenrauch, H., and Severinson Eklundh, K. (2005). D1.3.1 report on the evaluation methodology of multi-modal dialogue. Technical report, COGNIRON The Cognitive Robot Companion Integrated Project Information Society Technologies Priority, FP6-IST-002020.

Green, A., Hüttenrauch, H., Topp, E. A., and Eklundh, K. S. (2006a). Developing a Contextualized Multimodal Corpus for Human-Robot Interaction. In *Proceedings of the Fifth international conference on Language Resources and Evaluation LREC2006*.

Green, A. and Severinson Eklundh, K. (2003). Designing for Learnability in Human-Robot Communication. *IEEE Transactions on Industrial Electronics*, 50(4):644–650.

Green, A., Wrede, B., Severinson Eklundh, K., and Li, S. (2006b). Integrating Miscommunication Analysis in the Natural Language Interface Design for a Service Robot. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems 2006 (IROS'06)*, pages 4678–4683, Beijing, China.

Grice, J. P. (1975). Logic and conversation. In Cole, P. and Morgan, J. L., editors, *Syntax and Semantics*, volume 3: Speech Acts, pages 41–58. Academic Press, New York, NY.

Grosz, B. J. and Sidner, C. L. (1986). Attentions, Intentions, and the Structure of Discourse. *Computational Linguistics*, 12(3):175–204.

Gustafson, J. (2002). *Developing Multimodal Spoken Dialogue Systems – Empirical Studies of Spoken Human-Computer Interaction*. PhD thesis, KTH, Institutionen för talöverföring och musikakustik.

Gustafson, J., Bell, L., Boye, J., Lindström, A., and Wirén, M. (2004). The NICE Fairy-tale Game System. In *Proceedings of SIGDIAL'04 (5th SIGdial Workshop on Discourse and Dialog)*, Cambridge, MA. NAACL.

Gustafson, J. and Sjölander, K. (2002). Voice transformations for improving children's speech recognition in a publicly available dialogue system. In *Proceedings of International Conference On Spoken Language Processing (ICSLP)*, pages 297–300.

Haasch, A., Hohenner, S., Huewel, S., Kleinehagenbrock, M., Lang, S., Toptsis, I., Fink, G. A., Fritsch, J., Wrede, B., , and Sagerer, G. (2004). BIRON – The Bielefeld Robot Companion. In *Proceedings of ASER 2004 - 2nd International Workshop on Advances in Service Robots*, Stuttgart, Germany.

Hall, E. T. (1966). *The Hidden Dimension: Man's Use of Space in Public and Private*. The Bodley Head Ltd, London, UK.

Helbing, D. and Molnár, P. (1995). Social force model for pedestrian dynamics. *Physical Review E*, 51(5):4282–4286.

Hillier, B. and Hanson, J. (1984). *The Social Logic of Space*. Cambridge Press, Cambridge.

Holzapfel, H., T. Schaaf, H. E., Schaa, C., and Waibel, A. (2006). A robot learns to know people - First contacts of a Robot. In *Lecture Notes in Artificial Intelligence KI 2006,*. Springer, Bremen, Germany.

Hone, K. and Baber, C. (2001). Designing habitable dialogues for speech-based interaction with computers. *International Journal of Human Computer Studies*, 54(4):637–662.

Hone, K. S. and Graham, R. (2000). Towards a tool for the subjective assessment of speech system interfaces (SASSI). *Natural Language Engineering,*, 6(3/4):287–305.

Huang, H.-M. (2007). Autonomy Levels for Unmanned Systems (ALFUS) Framework: Safety and Application Issues. In *Proceedings of the Performance Metrics for Intelligent Systems (PerMIS) Workshop*, page August, Gaithersburg, MD, USA.

Huang, H.-M., Pavek, K., Novak, B., Albus, J., and Messina, E. (2005). A framework for autonomy levels for unmanned systems (alfus),. In *Proceedings of the AUVSI's Unmanned Systems North America 2005,*, Baltimore, Maryland, USA.

Hüttenrauch, H. (2007). *From HCI to HRI: Designing Interaction for a Service Robot*. PhD thesis, KTH Royal Institute of Technology. PhD Thesis.

Hüttenrauch, H., Green, A., Norman, M., Oestreicher, L., and Severinson Eklundh, K. (2004). Involving Users in the Design of a Mobile Office Robot. *Systems, Man and Cybernetics, Part C: Applications and reviews*, 34(2):113–124.

Hüttenrauch, H. and Norman, M. (2001). PocketCERO – mobile interfaces for service robots. In *Proceedings of Mobile HCI 2001: Third International Workshop on Human Computer Interaction with Mobile Devices*, Lille, France. IHM-HCI.

Hüttenrauch, H. and Severinson Eklundh, K. (2003). To Help or Not to Help a Service Robot. In *Proceedings of the 12th IEEE International Workshop on Robot and Human Interactive Communication RO-MAN'2003*, Millbrae CA, USA. IEEE.

Hüttenrauch, H., Severinson Eklundh, K., Green, A., and Topp, E. A. (2006a). Investigating Spatial Relationships in Human-Robot Interaction. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2006)*, pages 5052–5059.

Hüttenrauch, H., Severinson Eklundh, K., Green, A., Topp, E. A., and Christensen, H. I. (2006b). What's in the Gap? Interaction Transitions that make HRI work. In *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2006)*, Hatfield, UK.

Isendor, I. (1998). Mänsklig interaktion med autonom servicerobot. Technical report, KTH NADA, Royal Institute of Technology. Report no. TRITA-NA-E9841, IPLab-148.

Ishiguro, H. and Minato, T. (2005). Development of Androids for Studying Human-Robot Interaction. In *Proceedings of 36th International Symposium on Robotics*, volume TH3H1.

James, F., Rayner, M., and Hockey, B. A. (2000). 'Do that again': Evaluating spoken dialogue interfaces. Technical report, RIACS Technical Report #00.06.

Jones, J. L. (2006). Robots at the tipping point: the road to iRobot Roomba. *Robotics & Automation Magazine, IEEE*, 13(1):76–78.

Jönsson, A. and Dahlbäck, N. (2000). Distilling dialogues - A method using natural dialogue corpora for dialogue systems development. In *Proceedings of 6th Applied Natural Language Processing Conference*, pages 44–51, Seattle, WA, USA.

Kahn, P., Freier, N. G., Friedman, B., Severson, R. L., and Feldman, E. N. (2004). Social and Moral Relationships with Robotic others? *13th IEEE International Workshop on Robot and Human Interactive Communication, 2004. RO-MAN 2004*, pages 545–550.

Kahn, P. H., Friedman, B., Alexander, I. S., Freier, N., and Collett, S. (2005). The distant gardener: what conversations in the Telegarden reveal about human-telerobotic interaction. In *Proceedings of the 14th International Workshop on Robot and Human Interactive Communication (RO-MAN'05)*, pages 13–18.

Kahn, P. H., Friedman, B., Perez-Granados, D., and Freier, N. G. (2006). Robotic pets in the lives of preschool children. *Interaction Studies*, 7(3):405–436.

Kahn, Z. (1998). Attitudes Towards Intelligent Service Robots. Technical Report TRITA-NA-E98421 - IPLab-154, IPLab, NADA, Royal Institute of Technology.

Kaindl, H., Falb, J., and Bogdan, C. (2008). Multimodal Communication Involving Movements of a Robot. In *CHI '08 extended abstracts on Human factors in computing systems*, pages 3213–3218, New York, NY, USA. ACM.

Kanda, T., Hirano, T., Eaton, D., and Ishiguro, H. (2004). Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial. *Human Computer Interaction (Special issue on Human-Robot Interaction)*, 19(1-2):61–84.

Kanda, T., Ishiguro, H., Ono, T., Imai, M., and Mase, K. (2002a). Multi-robot Cooperation for Human-Robot Communication. In *IEEE International Workshop on Robot and Human Communication (Ro-Man2002)*, pages 271–276.

Kanda, T., Ishiguro, H., Ono, T., Imai, M., and Nakatsu, R. (2002b). Development and Evaluation of an Interactive Humanoid Robot "Robovie". In *IEEE International Conference on Robotics and Automation (ICRA 2002)*, pages 1848–1855.

Kanda, T., Sato, R., Saiwaki, N., and Ishiguro, H. (2007). A Two-Month Field Trial in an Elementary School for Long-Term Human–Robot Interaction. *IEEE Transactions on Robotics*, 23(5):962–971.

Kanto, K., Cheadle, M., Gambäck, B., Hansen, P., Kristiina Jokinen, H. K., and Rissanen, J. (2003). Multi-Session Group Scenarios for Speech Interface Design. In Stephanidis, C. and Jacko, J., editors, *Human-Computer Interaction: Theory and Practice (Part II)*, volume 2, pages 676–680. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Karsenty, L. (2001). Adapting verbal protocol methds to investigate speech systems use. *Applied Ergonomics*, 32:15–22.

Karsenty, L. (2002). Shifting the Design Philosophy of Spoken Natural Language Dialogue: From Invisible to Transparent Systems. *International Journal of Speech Technology*, 5(2):147–157.

Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Information Systems (TOIS)*, 2(1):26–41.

Kendon, A. (1990). *Conducting interaction - Patterns of behavior in focused encounters. Studies in interactional sociolinguistics*. Press syndicate of the University of Cambridge, Cambridge, NY, USA.

Kendon, A. (1997). Gesture. *Annual Review of Anthropology*, 26:1089–128.

Kipp, M. (2004). *Gesture Generation by Imitation: From Human Behaviour to Computer Character Animation*. Dissertion.com, Boca Raton, Florida.

Klingspor, V., Demiris, J., and Kaiser, M. (1997). Human-Robot-Communication and Machine Learning. *Applied Artificial Intelligence Journal*, 11(719–746).

Knudsen, M. W., Dykjær, L., and Bernsen, N. O. (2001). Surveys of Multimodal Data Resources, Annotation Schemes and Tools. In *Proceedings of the COCOSDA'2001 Workshop on Language Resources and Technology Evaluation - Technical, Global and Regional Perspectives*, pages 135–146, Aalborg, Denmark.

Koay, K., Sisbot, E., Syrdal, D., Walters, M., Dautenhahn, K., and Alami, R. (2007). Exploratory study of a robot approaching a person in the context of handing over an object. In *AAAI 2007 Spring Symposia - Technical Report SS-07-07*, pages 18–24.

Koide, Y., Kanda, T., Sumi, Y., Kogure, K., and Ishiguro, H. (2004). An Approach to Integrating an Interactive Guide Robot with Ubiquitous Sensors. In *Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004. (IROS 2004)*, volume 3, pages 2500–2505.

Kowtko, J., Isard, S., and Doherty, G. (1991). Conversational games within dialogue. In *Proceedings of the ESPRIT Workshop on Discourse Coherence*.

Kuipers, B. (2007). An Intellectual History of the Spatial Semantic Hierarchy. In Jefferies, M. and Yeap, A. W.-K., editors, *Robot and Cognitive Approaches to Spatial Mapping*. Springer Verlag.

Langton, S. R. H., Watt, R. J., and Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, 4(2):50–59.

Larsson, S. (2002). *Issue-based Dialogue Management. PhD Thesis, Goteborg University*. PhD thesis, Göteborg University.

Larsson, S. (2003). Generating feedback and sequencing moves in a dialogue system. In Freedman and Callaway, editors, *Natural Language Generation in Spoken and Written Dialogue - Papers from the 2003 AAAI Symposium.*, pages 79–84, Menlo Park, California. AAAI Press.

Lauria, S., Bugmann, G., Kyriacou, T., and Klein, E. (2002). Mobile robot programming using natural language. *Robotics and Autonomous Systems*, 38(3-4):171–181.

Lemon, O., Gruenstein, A., and Peters, S. (2002). Collaborative activities and multi-tasking in dialogue systems. *Traitement Automatique de Langues (TAL)*.

Levinson, S. C. (1983). *Pragmatics*. Cambridge Textbooks in Linguistics. Cambridge University Press.

Levinson, S. C. (1996). Language and space. *Annual Review of Anthropology*, 25:353–382.

Lewin, K. (1939). Field theory and experiment in social psychology: Concepts and methods. *The American Journal of Sociology*, 44(6):868–896.

Li, S. (2007). *Multi-modal Interaction Management for a Robot Companion*. PhD thesis, University of Bielefeld.

Lohse, M., Hanheide, M., Wrede, B., Walters, M. L., Koay, K. L., Syrdal, D. S., Green, A., Hüttenrauch, H., Dautenhahn, K., Sagerer, G., and Severinson Eklundh, K. (2008). Evaluating extrovert and introvert behaviour of a domestic robot - a video study. In *17th IEEE International Conference on Robot and Human Interactive Communication Ro-Man2008*, Munich, Germany.

Mahani, M. N. (2006). Towards resolving ambiguities in service robots' behavior. Master's thesis, KTH Royal Institute of Technology.

Malhotra, A. (1975). Design criteria for a knowledge-based English language system for management: an experimental analysis. Technical Report MAC TR-146, MIT.

Martinez-Garcia, E., Ohya, A., and Yuta, S. (2006). Guiding a group of people by a team of mobile robots. *International Journal of Vehicle Autonomous Systems*, 4(2-4):308 – 327.

Martinez-Garcia, E. A., Ohya, A., and Yuta, S. (2005). Crowding and Guiding Groups of Humans by Teams of Mobile Robots. In *IEEE Workshop on Advanced Robotics and its Social Impacts TAR*.

Martinovski, B. and Traum, D. (2003). The Error is the Clue: Breakdown in Human-Machine Interaction. In *proceedings of the ISCA tutorial and research workshop on Error handling in dialogue systems*, pages 11–16.

Martinson, E. and Brock, D. (2006). Auditory Perspective Taking. In *HRI '06: Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 345–346, New York, NY, USA. ACM.

Martinson, E. and Brock, D. (2007). Improving Human-Robot Interaction Through Adaptation to the Auditory Scene. In *HRI '07: Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 113–120, New York, NY, USA. ACM.

Massaro, D. W. (1998). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. MIT Press, Cambridge, MA.

Matsui, T., Asoh, H., and Asano, F. (1997). Map learning of the office conversant mobile robot, JIJO-2, by dialogue guided navigation. In *Proceedings of International Conference on Field and Service Robots (FSR'97)*, Canberra, Australia.

Matsui, T., Asoh, H., and Fry, J. (1999). Integrated Natural Spoken Dialogue System of Jijo-2 Mobile Robot for Office Services. In *Proceedings of AAAI-99*, Florida.

Maulsby, D., Greenberg, S., and Mander, R. (1993). Prototyping an Intelligent Agent through Wizard of Oz. In *INTERCHI'93*, pages 277 – 282. ACM.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago University Press, Chicago.

McTear, M. F. (2002). Spoken Dialogue Technology: Enabling the Conversational User Interface. *ACM Computing Surveys (CSUR)*, 34(1).

Mehrabian, A. (1967). Orientation behaviors and nonverbal attitude communication. *Journal of Communication*, 17:324 – 332.

Michalowski, M., Sabanovic, S., and Simmons, R. (2006). A Spatial Model of Engagement for a Social Robot. In *9th IEEE International Workshop on Advanced Motion Control*, pages 762 – 767.

Miller, R. B. (1968). Response time in man-computer conversational transactions. In *Fall Joint Comp. Conf. U.S.A.*, pages 267–277.

Moratz, R. and Ragni, M. (2008). Qualitative spatial reasoning about relative point position. *Journal Visual Languages & Computing*, 19(1):75–98.

Moratz, R., T.Tenbrink, Fischer, K., and Bateman, J. (2003). Spatial knowledge representation for human-robot interaction. In Freksa, C., Habel, C., and Wender, K. F., editors, *Spatial Cognition III*, pages 263–286, Berlin. Springer Verlag.

Nichols, J., Myers, B. A., Higgins, M., Hughes, J., Harris, T. K., Rosenfeld, R., and Litwack, K. (2003). Personal universal controllers: controlling complex appliances with guis and speech. In *CHI '03: CHI '03 extended abstracts on Human factors in computing systems*, pages 624–625, New York, NY, USA. ACM.

Nielsen, J. (1994). Heuristic Evaluation. In Nielsen, J. and Mack, R., editors, *Usability Inspection Methods*. John Wiley & Sons, New York, NY.

Norman, D. A. (1990). *The Design of Everyday Things*. MIT Press, Cambridge, MA.

Norman, D. A. (2005). *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books.

Oestreicher, L. (2002). Task Patterns for Human-Robot Interaction. In *Task Models and Diagrams for User Interface Design: Proceedings of the First International Workshop on Task Models and Diagram for User Interface Design - TAMODIA 2002*, pages 96–103, Bucharest, Romania.

Oestreicher, L., Hüttenrauch, H., and Severinson Eklundh, K. (1999). Where are you going little robot? – Prospects of Human Robot Interaction. In *CHI-99 Basic Research Symposium*.

Otero, N., Green, A., Nehaniv, C., Hüttenrauch, H., Syrdal, D., Dautenhahn, K., and Severinson Eklundh, K. (2007). Insights from Corpora of Embodied Interaction with Cognitive Service Robots. Technical Report 472, School of Computer Science, University of Hertfordshire.

Pacchierotti, E., Christensen, H., and Jensfelt, P. (2005). Embodied Social Interaction in Hallway Settings: a Pilot User Study. In *Proceedings of the 14th IEEE International Symposium on Robot and Human Interactive Communication RO-MAN 2005*, pages 164–171, Nashville, TN.

Pashler, H., Johnston, J. C., and Ruthruff, E. (2001). Attention and Performance. *Annual Review of Psychology*, 52:629–651. Academic Press.

Perzanowski, D., Brock, D., Adams, W., Bugajska, M., Schultz, A. C., Trafton, J. G., Blisard, S., and Skubic, M. (2003). Finding the FOO: a pilot study for a multimodal interface. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 4, pages 3218–3223.

Perzanowski, D., Schultz, A., Adams, W., Marsh, E., and Bugajska, M. (2001). Building a multimodal human-robot interface. *Intelligent Systems, IEEE [see also IEEE Expert]*, 16(1):16–21.

Pickering, M. J. and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27:169–225.

Power, R. (1979). The organisation of purposeful dialogues. *Linguistics*, 107-152.

Quek, F., McNeill, D., Bryll, R., an d H. Arslan, C. K., McCullough, K., Furuyama, N., and Ansari, R. (2000). Gesture, Speech, and Gaze Cues for Discourse Segmentation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 247 – 254, Hilton Head Island, South Carolina.

Quinderé, M., Lopes, L. S., and Teixeira, A. (2007). An information state dialogue manager for a mobile robot. In *Proceedings of Interspeech'2007*, pages 162–165, Antwerp, Belgium.

Renz, J. and Nebel, B. (1999). On the complexity of qualitative spatial reasoning: A maximal tractable fragment of the region connection calculus. *Artificial Intelligence*, 108(1-2):69–123.

Rosenfeld, R., Olsen, D., and Rudnicky, A. (2000). Universal Human-Machine Speech Interface: A White paper. Technical report, Carnegie Mellon University. CMU-CS-00-114.

Rosset, S., Bennacef, S., and Lamel, L. (1999). Design strategies for spoken language dialog systems. In *Proceedings of the European Conference on Speech Technology, EuroSpeech*, pages 1535–1538, Budapest.

Sack, R. (1986). *Human Territoriality: Its Theory and History*. Cambridge University Press, Cambridge.

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

Salter, T., Dautenhahn, K., and Boekhorst, R. (2006). Learning about natural human-robot interaction styles. *Robotics and Autonomous Systems*, 54(2):127–134.

Schegloff, E. (1998). Body torque. *Social Research*, 65(3):535–596.

Schegloff, E. A. (1979). Identification and recognition in telephone conversation openings. In Psathas, G., editor, *Everyday language studies in ethnomethodology*, pages 23–78. Irvington Publishers, New York.

Scheutz, M., Schermerhorn, P., Kramer, J., and Anderson, D. (2007). First steps toward natural human-like hri. *Autonomous Robots*, 22(4):411–423.

Schiel, F., Steininger, S., and Türk, U. (2002). The SmartKom Multimodal Corpus at BAS. In *Proceedings of Second International Conference on Language Resources and Evaluation LREC2000*, pages 200–206.

Schrage, M. (2004). Never Go to a Client Meeting without a Prototype. *IEEE Software*, 21(2):42–45.

Schulte, J., Rosenberg, C., and Thrun, S. (1999). Spontaneous short-term interaction with mobile robots in public places. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'99)*.

Searle, J. R. (1969). *Speech Acts - an Essay in The Philosophy of Language*. Cambridge University Press, Cambridge.

Severinson Eklundh, K. (1983). The Notion of Language game – A Natural Unit of Dialogue and Discourse. Technical Report SICS 5, University of Linköping.

Severinson Eklundh, K., Espmark, E., Green, A., Hüttenrauch, H., Norman, M., and Oestreicher, L. (2001). Användaranpassad utformning av en ”fetch-and-carry”-robot för funktionshindrade i arbetslivet. Technical Report TRITA-NA-P0110, IPLab Report IPLab-186., KTH NADA.

Severinson Eklundh, K., Green, A., and Hüttenrauch, H. (2003). Social and collaborative aspects of interaction with a service robot. *Robotics and Autonomous Systems*, 42(3–4):223–234. Special issue on Socially Interactive Robots.

Shiomi, M., Kanda, T., Koizumi, S., Ishiguro, H., and Hagita, N. (2007). Group Attention Control for Communication Robots with Wizard of OZ Approach. In *HRI '07: Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 121–128, New York, NY, USA. ACM.

Shiwa, T., Kanda, T., Imai, M., Ishiguro, H., and Hagita, N. (2008). How quickly should communication robots respond? In *HRI '08: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 153–160, New York, NY, USA. ACM.

Shneiderman, B. and Plaisant, C. (2004). *Designing the User Interface: Strategies for Effective Human-Computer Interaction (4th Edition)*. Pearson Addison Wesley.

Sidner, C. L., Kidd, C. D., Lee, C., and Lesh, N. (2004). Where to Look: a Study of Human-Robot Engagement. In *IUI'04: Proceedings of the 9th international conference on Intelligent User Interfaces*, pages 78–84, New York, NY, USA. ACM Press.

Sidner, C. L., Lee, C., Kidd, C. D., Lesh, N., and Rich, C. (2005). Explorations in engagement for humans and robots. *Artificial Intelligence*, 166:140–164.

Siegwart, R., Arras, K. O., Bouabdallah, S., Burnier, D., Froidevaux, G., Greppin, X., Jensen, B., Lorotte, A., Mayor, L., Meisser, M., Philippsen, R., Piguet, R., Ramel, G., Terrien, G., and Tomatis, N. (2003). Robox at Expo.02: A Large-Scale Installation of Personal Robots. *Robotics and Autonomous Systems*, 42(3-4):203–222.

Sinclair, J. M. and Coulthard, R. M. (1975). Towards an analysis of discourse: The English used by teachers and pupils. *Computational Intelligence*, 1(1):1–10. Oxford University Press, London.

Skantze, G. (2007). *Error Handling in Spoken Dialogue Systems - Managing Uncertainty, Grounding and Miscommunication*. PhD thesis, KTH Royal Institute of Technology, Stockholm, Sweden.

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

Syrdal, D. S., Dautenhahn, K., Woods, S. N., Walters, M. L., and Koay, K. L. (2006). 'Doing the Right Thing Wrong' - Personality and Tolerance to Uncomfortable Robot Approaches. In *Proc. 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06)*, pages 183–188.

Tåqvist, H. (1999). Fetch-and-carry robot for assistance to humans. a navigation system and a human following behaviour. Master's Thesis TRITA-NA-E9955., Royal Institute of Technology, NADA.

Tasaki, T., Komatani, K., Ogata, T., and Okuno, H. (2005). Spatially Mapping of Friendliness for Human-Robot Interaction. *Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 1277–1282.

Tellex, S. and Roy, D. (2006). Spatial Routines for a Simulated Speech-Controlled Vehicle. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, HRI*, Salt Lake City, Utah, USA. ACM 2006.

Tenbrink, T. (2003). Communicative Aspects of Human-Robot Interaction. In Metslang, H. and Rannut, M., editors, *Languages in development*. Lincom Europa.

Tenbrink, T., Fischer, K., and Moratz, R. (2002). Spatial Strategies in Human-Robot Communication. *KI 4/02*. Themenheft Spatial Cognition Christian Freksa (ed.) arenDTaP Verlag.

Thorisson, K. (1997). A Mind Model for Multimodal Communicative Creatures & Humanoids. *International Journal of Applied Artficial Intelligence*, 13(4-5):449–486. Special Issue on Animated Interface Agents.

Thrun, S., Bennewitz, M., Burgard, W., Cremers, A., Dellaert, F., Fox, D., Hähnel, D., Rosenberg, C., Roy, N., Schulte, J., and Schulz, D. (1999). MINERVA: A second generation mobile tour-guide robot. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'99)*.

Tomko, S., Harris, T. K., Toth, A., Sanders, J., Rudnicky, A., and Rosenfeld, R. (2005). Towards efficient human machine speech communication: The speech graffiti project. *ACM Trans. Speech Lang. Process.*, 2(1):2.

Topp, E. A. and Christensen, H. I. (2005). Tracking for Following and Passing Persons. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2005)*, Edmonton, Alberta.

Topp, E. A., Hüttenrauch, H., Christensen, H., and Eklundh, K. S. (2006). Acquiring a Shared Environment Representation. In *In Proceedings of HRI2006 1st annual conference on Human-Robot Interaction*, Salt Lake City, UT, USA. ACM.

Torrance, M. C. (1994). Natural Communication with Robots. Master's thesis, MIT Department of Electrical Engineering and Computer Science.

Torrey, C., Powers, A., Fussell, S. R., and Kiesler, S. (2007). Exploring Adaptive Dialogue Based on a Robot's Awareness of Human Gaze and Task Progress. In *HRI '07: Proceeding of the ACM/IEEE international conference on Human-robot interaction*, pages 247–254, New York, NY, USA. ACM Press.

Trafton, J. G., Cassimatis, N., Bugajska, M. D., Brock, D. P., Mintz, F. E., and Schultz, A. C. (2005). Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans.*, 35(4):460–470.

Trafton, J. G., Schultz, A. C., Cassimatis, N. L., Hiatt, L. M., Perzanowski, D., Brock, D. P., Bugajska, M. D., and Adams, W. (2006). Cognition and Multi-Agent Interactions From Cognitive Modeling to Social Simulation. In Sun,

R., editor, *Communicating and Collaborating with Robotic Agents*, chapter 10, pages 252–278. Cambridge University Press.

Traum, D. and Dillenbourg, P. (1996). Miscommunication in Multi-modal Collaboration. In *In working notes of the AAAI Workshop on Detecting, Repairing, And Preventing Human–Machine Miscommunication*, pages 37–46.

Traum, D. R. (1994). A computational theory of grounding in natural language conversation. Technical Report TR545, University of Rochester, Computer Science Department.

Traum, D. R. (1996). Conversational Agency: The Trains-93 Dialogue Manager. In *Proceedings of the Twente Workshop in Language Technology: Dialogue Management in Natural Language Systems (TWLT11)*, pages 1–11.

Traum, D. R. and Allen, J. F. (1994). Discourse Obligations in Dialogue Processing. In Pustejovsky, J., editor, *Proceedings of the Thirty-Second Meeting of the Association for Computational Linguistics*, pages 1–8, San Francisco.

Traum, D. R. and Heeman, P. A. (1996). Utterance Units and Grounding in Spoken Dialogue. In *Proc. ICSLP '96*, volume 3, pages 1884–1887, Philadelphia, PA.

Tschichold-Gürman, S. J., Vestli, G., and Schweitzer, G. (1999). Operating Experiences with the Service Robot MOPS. In *Proceedings of the 3rd European Workshop on Advanced Mobile Robots*, Zurich, Switzerland.

Tversky, B. and Lee, P. U. (1998). How space structures language. In *Spatial Cognition: An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*, volume 1404/1998 of *Lecture Notes in Computer Science*. Springer, Berlin / Heidelberg.

Van Den Haak, M., De Jong, M., and Schellens, P. (2003). Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & Information Technology*, 22(5):339–351.

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 (ACL-97), pages 271–280, Madrid, Spain.

Walters, M., Woods, S., Koay, K., and Dautenhahn, K. (2005a). Practical and methodological challenges in designing and conducting interaction studies with human subjects. In *Proceeding of AISB Symposium on Robot Companions Hard Problems and Open Challenges in Human-Robot Interaction*, pages 110–120, Hertfordshire UK.

Walters, M. L., Dautenhahn, K., Koay, K. L., Kaouri, C., te Boekhorst, R., Nehaniv, C. L., Werry, I., and Lee, D. (2005b). Close encounters: Spatial distances between people and a robot of mechanistic appearance. In *Proceedings of IEEE-RAS International Conference on Humanoid Robots (Humanoids2005)*, pages 450–455, Tsukuba, Japan.

Walters, M. L., Koay, K. L., Woods, S. N., Syrdal, D. S., and Dautenhahn, K. (2007). Robot to human approaches: Comfortable distances and preferences. In *Proceedings of the AAAI Spring Symposium on Multidisciplinary Collaboration for Socially Assistive Robotics, (AAAI SS07-2007).*, Palo Alto, California.

Walters, M. L., Syrdal, D. S., Dautenhahn, K., te Boekhorst, R., and Koay, K. L. (2008). Avoiding the uncanny valley – robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Journal of Autonomous Robots*, 24(2):159–178.

Widlok, T. (1999). Mapping spatial and social permeability. *Current Anthropology*, 40(3):392–400.

Wijk, O. and Christensen, H. I. (1999). Localization and navigation of a mobile robot using natural point landmarks extracted from sonar data. *Robotics and Autonomous System*. special issue.

Wilson, H. R., Wilkinson, F., Lin, L.-M., and Castillo, M. (2000). Perception of head orientation. *Vision Research*, 40(4):459–472.

Wolf, J. C. and Bugmann, G. (2005). Multimodal Corpus Collection for the Design of User-Programmable Robots. In *TAROS 2005 Towards Autonomous Robotic Systems Incorporating the Autumn Biro-Net Symposium*.

Yamaoka, F., Kanda, T., Ishiguro, H., and Hagita, N. (2006). How Contingent Should a Communication Robot Be? In *HRI '06: Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 313–320, New York, NY, USA. ACM.

Yankelovich, N. (1996). How Do Users Know What to Say? *ACM Interactions*, 3(6).

Zelek, J. S. (1997). Human-Robot Interaction with inimal Spanning Natural Language Template for Autonomous and Tele-Operated Control. *Proceedings of the 1997 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '97)*, 1:299–305 vol.1.

Zender, H., Jensfelt, P., and Kruijff, G.-J. M. (2007). Human- and Situation-Aware People Following. In *16th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2007)*, Jeju Island, Korea.

Zoltan-Ford, E. (1991). How to Get People to Say and Type What Computers Can Understand. *International Journal of Man-Machine Studies*, 34:527–547.