

Affordance-Based Task Communication Methods for Astronaut-Robot Cooperation

Seppo S. Heikkilä



Affordance-Based Task Communication Methods for Astronaut-Robot Cooperation

Seppo S. Heikkilä

Doctoral dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the School of Electrical Engineering for public examination and debate in Auditorium AS1 at the Aalto University School of Electrical Engineering (Espoo, Finland) on the 18th of November 2011 at 12 noon.

Aalto University
School of Electrical Engineering
Department of Automation and Systems Technology
Centre of Excellence in Generic Intelligent Machines Research

Supervisor

Prof. Aarne Halme

Preliminary examiners

Prof. Mark Neerincx, Delft University of Technology, The Netherlands

Prof. Miguel Salichs, Carlos III University of Madrid, Spain

Opponents

Dr. Matti Anttila, Space Systems Finland, Finland

Prof. Mark Neerincx, Delft University of Technology, The Netherlands

Aalto University publication series

DOCTORAL DISSERTATIONS 102/2011

© Seppo S. Heikkilä

ISBN 978-952-60-4330-2 (pdf)

ISBN 978-952-60-4329-6 (printed)

ISSN-L 1799-4934

ISSN 1799-4942 (pdf)

ISSN 1799-4934 (printed)

Aalto Print

Helsinki 2011

Finland

The dissertation can be read at <http://lib.tkk.fi/Diss/>



Author

Seppo Heikkilä

Name of the doctoral dissertation

Affordance-Based Task Communication Methods for Astronaut-Robot Cooperation

Publisher School of Electrical Engineering**Unit** Department of Automation and Systems Technology**Series** Aalto University publication series DOCTORAL DISSERTATIONS 102/2011**Field of research** Automation technology**Manuscript submitted** 1 June 2011**Manuscript revised** 7 October 2011**Date of the defence** 18 November 2011**Language** English **Monograph** **Article dissertation (summary + original articles)****Abstract**

The problem with current human-robot task communication is that robots cannot understand complex human speech utterances, while humans cannot efficiently use the fixed task request utterances required by robots. Nonetheless, future planetary exploration missions are expected to require astronauts on extra-vehicular activities to communicate task requests to robot assistants with speech- and gesture-type user interfaces that can be easily embedded in their space suits.

The solution proposed in this thesis is indirect task communication based on the human-like ability to utilise object-action relationships in task communication. Conventional task communication methods, in which all task parameters need to be communicated explicitly, are evaluated against task communication methods where affordances, i.e. action possibilities, are used to complete task communication.

These so-called affordance-based task communication methods are evaluated by means of four user experiments: two performed with a fully autonomous centauroid robot in a planetary exploration work context and two with a simulated robot in a lander assembly work context. The first two experiments are performed in unambiguous work environments, where each object is associated with only one action and vice versa, while the last two experiments are performed in ambiguous work environments, where each object and action is normally associated with several actions and objects, respectively.

The user experiments show that affordance-based task communication methods can be used to decrease both the human workload and task communication times in a planetary exploration work context. Furthermore, affordance-based task communication methods are found to be preferred over conventional task communication methods.

The affordance-based task communication methods derived can be applied to facilitate any human-robot task communication that includes a priori known or recurring task sequences. In this thesis, the feasibility of the approach was demonstrated for frame-based dialogue managers, which are widely used in robotics.

Keywords human-robot interaction, robotic astronaut assistant, task communication, affordance**ISBN (printed)** 978-952-60-4329-6**ISBN (pdf)** 978-952-60-4330-2**ISSN-L** 1799-4934**ISSN (printed)** 1799-4934**ISSN (pdf)** 1799-4942**Location of publisher** Espoo**Location of printing** Espoo**Year** 2011**Pages** 176**The dissertation can be read at** <http://lib.tkk.fi/Diss/>

Tekijä

Seppo Heikkilä

Väitöskirjan nimi

Affordanssi-pohjaisia tehtävienkommunikointimenetelmiä astronautti-robotti yhteistyöhön

Julkaisija Sähkötekniikan korkeakoulu**Yksikkö** Automaatio- ja systeemitekniikan laitos**Sarja** Aalto University publication series DOCTORAL DISSERTATIONS 102/2011**Tutkimusala** Automaatiotekniikka**Käsikirjoituksen pvm** 01.06.2011**Korjatun käsikirjoituksen pvm** 07.10.2011**Väitöspäivä** 18.11.2011**Kieli** Englanti **Monografia** **Yhdistelmäväitöskirja (yhteenvedo-osa + erillisartikkelit)****Tiivistelmä**

Ongelmana nykyisessä ihmisen ja robotin välisessä kommunikoinnissa on se, että robotti ei pysty ymmärtämään monimutkaista ihmisen puhetta, kun taas ihminen ei pysty tehokkaasti hyödyntämään määrämuotoisia lausahduksia, joita robotti pystyisi helposti käsittelemään. Tutkimusmatkat planeettojen pinnoille vaativat kuitenkin tulevaisuudessa avaruuskävelyillä olevia astronautteja viestimään tehtäviä avustaville roboteille puhe- ja elepohjaisilla käyttöliittymillä, joita voidaan käyttää helposti osana heidän avaruuspukujaan.

Tässä väitöskirjassa esitettävä ratkaisu on epäsuora tehtävien kommunikointi, joka hyödyntää robotissa ihmismäistä kykyä ymmärtää toimintojen ja kohteiden välisiä riippuvuussuhteita tehtäviä viestittäessä. Perinteistä tehtävien kommunikointia, jossa kaikki tehtävän parametrit viestitetään suoraan, verrataan affordanssi-pohjaiseen tehtävien kommunikointiin, jossa tietoa toimintamahdollisuuksista hyödynnetään tehtävien viestinnän loppuunsaattamisessa.

Tätä affordanssi-pohjaista tehtävien kommunikointia arvioidaan neljän käyttäjätestin avulla. Kaksi ensimmäistä käyttäjätestiä suoritetaan täysin autonomisella kentauri-tyyppisellä robotilla planeetantutkimuskontekstissa, jossa jokainen toiminto liittyy vain yhteen kohteeseen ja päinvastoin. Kaksi viimeistä käyttäjätestiä suoritetaan simuloidulla robotilla laskeutujan kokoamiskontekstissa, jossa jokainen toiminto on pääsääntöisesti yhdistettynä useaan kohteeseen ja päinvastoin.

Käyttäjätestit osoittavat, että affordanssi-pohjaista tehtävien kommunikointia voidaan käyttää keventämään ihmisen työkuormitusta ja tehtävien viestintäaikoja planeettatutkimus-tyyppisissä työkonteksteissa. Lisäksi affordanssi-pohjaiset menetelmät osoittautuvat suosituimmiksi tehtävien viestimässä kuin vertailut perinteiset menetelmät.

Esitettyjä affordanssi-pohjaisia kommunikointimenetelmiä voidaan hyödyntää teoriassa sellaisenaan helpottamaan mitä tahansa ihmisen ja robotin välistä tehtävien viestintää, joka sisältää ennalta tunnettuja tai toistuvia tehtäväsarjoja. Tässä väitöskirjassa näiden menetelmien toteutuskelpoisuus todistettiin lomakepohjaisille dialogimanagereille, joita käytetään laajasti robotiikassa.

Avainsanat ihminen-robotti vuorovaikutus, astronauttia avustava robotti, tehtävien kommunikointi, affordanssi

ISBN (painettu) 978-952-60-4329-6**ISBN (pdf)** 978-952-60-4330-2**ISSN-L** 1799-4934**ISSN (painettu)** 1799-4934**ISSN (pdf)** 1799-4942**Julkaisupaikka** Espoo**Painopaikka** Espoo**Vuosi** 2011**Sivumäärä** 176**Luettavissa verkossa osoitteessa** <http://lib.tkk.fi/Diss/>

Preface

This thesis for the degree of Doctor of Science in Technology is the result of the work done in the Department of Automation Technology at Aalto University, and in the European Space Research and Technology Centre (ESTEC) at the European Space Agency (ESA) during the years 2007-2011. The research was mainly carried out in a project called SpacePartner (ESTEC/Contract No. 21139/07/NL/EM), which was funded half by the Finnish Centre of Excellence in Generic Intelligent Machines Research (GIM) and half by ESA's Network Partnering Initiative (NPI).

I would like to acknowledge the personal support of the Jenny and Antti Wihuri Foundation (Jenny ja Antti Wihurin rahasto), the KAUTE foundation (KAUTE säätiö), and the Walter Ahlström Foundation (Walter Ahlströmin säätiö). The research grants allowed me to fully and unconditionally focus on this thesis work.

I wish to express my warmest gratitude to Professor Aarne Halme, my supervisor, for all his guidance and support throughout the years of this thesis and for his help in arranging funding for the research. It has been a great pleasure to work and explore ideas under your supervision in an encouraging and open-minded work environment.

I also wish to express my gratitude towards Dr. Andre Schiele and Frederic Didot for welcoming me to ESTEC in the Netherlands and for helping me to direct my research. You always managed to find time in your busy schedules to help and give answers whenever needed. Many thanks also go to Gianfranco Visentin and other members of ESA's Automation and Robotics group for their support during my stay in ESTEC.

Also, I would like to thank my preliminary examiners Professor Mark Neerincx and Professor Miguel A. Salichs for their reviews and valuable suggestions on the thesis.

Many people at the Automation Technology Laboratory provided me with valuable help and assistance on various matters during the thesis work. Thank you all for the friendly and inspiring work atmosphere. I would especially like to thank: Dr. Jari Saarinen, Antti Maula, Matthieu Myrsky, and Johan Grönholm for assisting with the GIMnet/MaCI software framework; Dr. Mikko Elomaa, Sami Kielosto and Petri Hyötylä for helping with electronics; Tapio Leppänen for helping with the mechanics; Professor Jussi Suomela, and Tomi Ylikorpi for supporting and promoting the SpacePartner project; Janne Paanajärvi for providing guidance with algorithms; Dr. Tapio Taipalus for reviewing parts of this thesis work; and Iris Mielonen, Erich Halbach, and Robert Guinness for assisting with the nuances of English language. Thanks are likewise due to everyone who helped with the WorkPartner robot, especially Dr. Sami Terho, Dr. Mikko Heikkilä, and Dr. Ilkka Leppänen.

A big thank you goes to Paavo Heiskanen, Melak Zebenay, and Mikael Persson for the good and dedicated work in your M.Sc. theses that underpin this research. I would also like to thank all the B.Sc. thesis and automation technology project workers involved in this thesis, and the 74 volunteers who took part in the thesis' user experiments. This thesis would not have been possible without your interest in this research.

Finally, I want to express gratitude to my family for their support for my studies, especially to my father, mother and brother: *paljon kiitoksia*. My deepest thanks to Dr. Cynthia Jimenez Monroy: *muchas gracias*, my lovely space partner, for always being there and for giving me all the support one could ever wish for.

Espoo, October 2011

Seppo S. Heikkilä

Contents

Preface	7
Contents	9
List of Abbreviations	13
List of Figures	15
List of Tables	17
1 Introduction	19
1.1 Background and motivation	19
1.2 Case study: SpacePartner project	21
1.3 Problem formulation	23
1.4 Research hypothesis and methodology	25
1.5 Main contributions of the dissertation	26
1.6 Author's contribution	27
1.7 Declaration of previous work	27
1.8 Thesis outline	28
2 State of the Art Regarding Robotic Assitants	29
2.1 Terminology	29
2.1.1 Action, task and mission	29
2.1.2 Collaboration, cooperation and coordination	31
2.1.3 Interface, interaction and communication	32
2.1.4 Natural and intuitive human-robot interaction	33
2.2 Robotic assistants	34
2.2.1 Research in astronaut-robot cooperation	36
2.2.2 Research in human-robot cooperation	40
2.3 Human-robot interfaces	42

2.3.1	Gesture interfaces	42
2.3.2	Visual displays	45
2.3.3	Speech interfaces	46
2.3.4	Haptic interfaces	47
2.4	Human inspired human-robot interaction	48
2.4.1	Design of human-robot interaction systems	49
2.4.2	Natural and indirect human-robot communication	50
2.4.3	Peer-to-peer dialogue	51
2.4.4	Perspective taking	53
2.4.5	Common ground	55
2.4.6	Deictic terms and gestures	57
2.5	Affordances - action possibilities	59
2.5.1	Concept of affordances	60
2.5.2	Affordances in user interface research	62
2.5.3	Affordances in robotics research	63
2.6	Conclusion	64
3	Requirements for Robotic Astronaut Assistants	65
3.1	EVA astronaut mission scenarios	65
3.2	Mission scenario breakdown	70
3.3	Robotic astronaut assistant requirements	72
3.3.1	WorkPartner robot	73
3.3.2	WorkPartner readiness	75
3.4	Conclusion	76
4	Unambiguous Task Communication Using Affordances	79
4.1	Geological exploration experiment	79
4.1.1	Method	80
4.1.2	Results	93
4.1.3	Discussion	95
4.2	Object manipulation experiment	97

4.2.1	Method	97
4.2.2	Results	106
4.2.3	Discussion	108
5	Ambiguous Task Communication Using Affordances	111
5.1	Task request prediction	111
5.1.1	Prediction with affordances	113
5.2	Predictive dialogue experiment	114
5.2.1	Method	114
5.2.2	Results	126
5.2.3	Discussion	129
5.3	Automatic execution experiment	131
5.3.1	Method	132
5.3.2	Results	139
5.3.3	Discussion	140
6	Conclusion	143
7	Future work	147
	References	149
	Appendices	167
A	Usage Examples of Control Development Methodology	169
B	Software Architectures of the User Experiments	173
B.1	GIM/MaCI software library	173
B.2	User experiments with unambiguous task communication	174
B.3	User experiments with ambiguous task communication	174

List of Abbreviations

API	Application Programming Interface
AR	Augmented Reality
ARC	Ames Research Center
ASRO	Astronaut-Rover
AV	Augmented Virtuality
BWG	Bag of Words in a Graph
CDM	Control Development Methodology
DOF	Degree Of Freedom
EGP	Eurobot Ground Prototype
ERA	EVA Robotic Assistant
ESA	European Space Agency
ESAS	Exploration Systems Architecture Study
EVA	Extra-Vehicular Activity
GIM	General Intelligent Machines
GUI	Graphical User Interface
HMM	Hidden Markov Model
HMMP	Human Mars Mission Project
HRI	Human-Robot Interaction
ISS	International Space Station

JSC	Johnson Space Center
LAN	Local Area Network
LED	Light-Emitting Diode
LEO	Lunar Exploration Objectives
LSRM	Lunar Surface Reference Mission
MaCI	Machine Control Interface
MISA	Mixed-Initiative Sliding Autonomy
MR	Mixed Reality
MSRM	Mars Surface Reference Mission
NASA	National Aeronautics and Space Administration
NLP	Natural Language Processing
NPI	Network Partnering Initiative
PDA	Personal Digital Assistant
SA	Situation Awareness
SISA	System-Initiative Sliding-Autonomy
SLAM	Simultaneous Localisation and Mapping
SPA	Sequence Prediction Algorithm
TLX	Task Load Index
TOF	Time Of Flight
VR	Virtual Reality
WDA	Work Domain Analysis

List of Figures

1.1	WorkPartner was used as a robotic astronaut assistant testbed . . .	21
1.2	The SpacePartner project was evaluated with user experiments . . .	22
2.1	Activity classes: mission, task, and subtask	30
2.2	Coordination, cooperation and collaboration	32
2.3	The astronaut and the Marsokhod cooperated in the ASRO project	37
2.4	The ERA project's robotic astronaut assistant	38
2.5	NASA's Robonaut performing planetary exploration tasks	39
2.6	ESA's EGP robot for human-interactive on-surface operations . . .	40
2.7	Virtuality continuum from real to virtual environments	45
2.8	General user interface module diagram	48
2.9	The peer-to-peer dialogue communication method	53
2.10	The perspective-taking communication method	55
2.11	Establishing a common ground communication method	57
2.12	The deictic references communication method	59
2.13	Example of indirect human-robot task communication	60
3.1	Capability requirements for a robotic astronaut assistant	74
4.1	Test area of the geological exploration user experiment	81
4.2	The WorkPartner robot worked as an astronaut assistant robot . .	81
4.3	The restraining outfit used to mimic an astronaut's space suit . . .	82
4.4	Measurement unit and rocks used in the experiment	83
4.5	Astronaut-robot cooperation system diagram	84
4.6	Dialogue structures of the task communication methods	87
4.7	Diagram of the dialogue manager subsystems	88
4.8	Complete progress of the experiment for one of the participants . .	91
4.9	Boxplot of the NASA-TLX workload in the first experiment	94
4.10	Boxplot of task communication method choices	95

4.11	Physical setup used in the second user experiment	98
4.12	Dialogue structures of the examined task communication methods .	101
4.13	Boxplot of the NASA-TLX workload in the second experiment . . .	107
4.14	Boxplot of task communication times in the second experiment . . .	108
4.15	Boxplot of task communication method utilisation	109
5.1	Correct, incorrect and non-possible task request predictions.	115
5.2	Physical setup used in the third user experiment	116
5.3	Screen shots from the video-based robot simulator	119
5.4	Twenty pictures depicting the tasks performed	121
5.5	Dialogue structures in the third user experiment	122
5.6	Boxplot of the NASA-TLX workload in the third experiment	126
5.7	Boxplot of test round execution times in the third experiment	127
5.8	Boxplot of communication method utilisation percentages	128
5.9	Boxplot of other than explicit task requests	129
5.10	List of 40 tasks performed by the participants	133
5.11	Dialogue structures in the fourth user experiment	134
5.12	Boxplot of other than explicit task requests	139
B.1	Software architecture used with the real robot	175
B.2	Software architecture used with the simulated robot	176

List of Tables

2.1	Comparison of motion capture technologies	43
3.1	EVA astronaut planetary surface activities	68
3.2	List of CDM tasks used during the missions	71
3.3	List of all subtasks used in the CDM tasks	72
3.4	WorkPartner readiness for astronaut assistance	75
4.1	All tasks and task requests in the first user experiment	86
4.2	Communication dialogues used in the first user experiment	90
4.3	All possible tasks in the second user experiment	100
4.4	Communication dialogues used in the second user experiment	103
5.1	List of all 65 possible tasks in the third user experiment	120
5.2	Example of task communication dialogue in the third experiment	124
5.3	Extracts from the participants' communication dialogues	138
5.4	Pros and cons of the task communication methods	140
A.1	Astronaut-robot LAN setup scenario tasks	169
A.2	Transport task decomposition to subtasks	170

1 Introduction

1.1 Background and motivation

The next manned missions to the surfaces of the Moon and Mars will be longer and more complex than any previous human spaceflights. It is expected that there will be a significant increase in the number of robotic assistants working with astronauts, and in the number of tasks astronauts are expected to perform, often without any assistance from ground control. This means that the astronauts' workload is also expected to increase significantly, since communicating tasks to robots is still cumbersome, especially when compared with the efficiency of human communication.

One way to cope with this increased complexity is to develop communication methods that can support human cognitive processes, i.e. are based on the way people naturally interact and process information. The way we naturally communicate with people in the real world could thus provide useful insights for better communication with robots in the future.

In fact, natural human-robot interaction, defined as a human-human type of interaction in the real world [105], has already been frequently mentioned as a desirable key element for future manned planetary missions to the surfaces of the Moon and Mars [36, 31]. A few human-inspired human-robot interaction methods such as peer-to-peer dialogue [35] and perspective-taking [134] have already been successfully shown to have potential for astronaut-robot task communication.

Peer-to-peer dialogue enables robots to be human peers, as they can use the humans as a resource by asking questions while executing tasks, just as people do. This was found to be of particular assistance to humans in understanding the problems encountered by the robots [35]. Alternatively, perspective-taking enables the robot

to reason and simulate the world from the perspective of others, which increases human flexibility in describing spatial locations [134].

One significant unsolved problem in human-robot interaction is how to have humans efficiently communicate task requests to robots. Task requests are defined here as consisting of at least the parameters of action and target object, which can be considered the minimum of parameters needed to define a proper task [90]. The underlying problem is that unconstrained human-to-human type of communication is so complex that it cannot be fully understood in practice by any robot in the near future. On the other hand, humans, on average, essentially need to communicate in this versatile and flexible manner and cannot be, for example, expected to learn and remember the dozens of fixed communication utterances that a robot could easily interpret [42, 73, 15, 65]. The mere use of synonyms, for example, cannot solve this problem because there are so many possible synonyms being used and their meanings overlap even in very restricted contexts [42].

This task request problem can be also approached by examining how humans communicate with each other, for instance, in special cases with additional communication constraints, somewhat similar to the ones imposed by robotic assistants. Such situations could be adult-child communication or guide-tourist communication, where shared communication abilities intersect only partially. It is known that in these cases people tend to use very low-level language if they do not expect the other person to correctly understand what is being said [132].

The same preference for using lower level language has already been identified when communicating with robots. Essentially, humans prefer to communicate with the robot on a level at which they think the robot will correctly understand them [132]. In most cases, this means simple utterances that do not leave any margin for misinterpretation. This reflects the basic requirement of human-robot task communication, which is that the task request utterances used need to be usable both for the

human and the robot. This fundamental requirement is also the starting point for the task communication method presented in this thesis.

1.2 Case study: SpacePartner project

Most of the research reported in this thesis was carried out within a research project called SpacePartner. The SpacePartner project was a Ph.D. project active between 2008 and 2011, co-sponsored by the European Space Agency (ESA) and Aalto University. The project was initiated under the ESA Network Partnering Initiative (NPI) program, whose goal is to increase interaction between the ESA and European universities. The ESA NPI program also aims to improve space research through spin-ins from advanced non-space projects. In this case, the spin-in is the use of Aalto University's WorkPartner service robot, shown in Figure 1.1, to develop astronaut-robot cooperative task definition and execution capabilities.

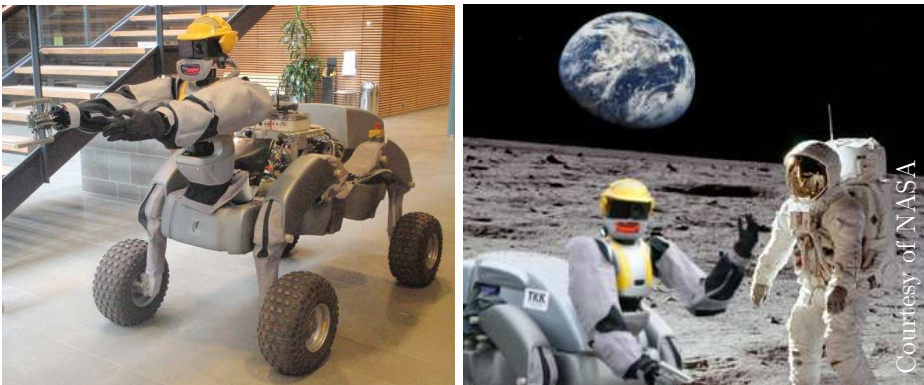


Figure 1.1: Aalto University's WorkPartner robot (left image) is used as an astronaut-robot cooperation test platform (right image - artistic impression).

The idea of the SpacePartner project was to focus on astronaut-robot interface development and on efficient information sharing between astronaut and robot. The

ready-to-use WorkPartner service robot made it possible to focus on core astronaut-robot interaction research problems, instead of working with robotic assistant platform development.

The SpacePartner project was demonstrated with several user experiments. In total, 28 participants from Aalto University took part in two user experiments conducted with the actual, fully autonomous WorkPartner robot. The first experiment, shown on the left in Figure 1.2, consisted of dealing with emerging problems when working next to a planetary lander facility. The second experiment, shown on the right in Figure 1.2, consisted of a simulated geological exploration mission where participants were requested to analyse rocks and set up measurement units. Through field tests, both of these experiments demonstrated the feasibility of the proposed task communication for human-robot cooperation. These experiments are presented in more detail in Chapter 4 of this thesis.



Figure 1.2: The SpacePartner project user experiments dealt with solving emerging problems (left image) and with performing geological exploration (right image).

Furthermore, 34 participants took part in two user experiments done with a simulated WorkPartner robot. These experiments aimed at extending the research to more complicated mission scenarios that could not be reasonably implemented with the actual robot. In these experiments, the participants acted as astronauts who

perform tasks with the WorkPartner robot on Mars in order to accomplish a lander setup mission. These experiments are presented in more detail in Chapter 5 of this thesis.

1.3 Problem formulation

The core problem addressed in this thesis arises from the fact that current human-robot task communication is inefficient in terms of human workload and task communication time. This is largely owing to the fact that humans and robots do not communicate tasks in the same way. Humans are not able to request tasks with the strictly defined communication utterances required by robots, and the robots are not able to understand complex natural human communication [42, 73, 65].

Based on this identified task communication problem, we can define the main research question to be:

“How could the human task communication workload and the task communication times be decreased, as compared with the current conventional task communication methods, when a human communicates tasks to a robot in the field by using mostly recognisable task sequences?”

Task communication workload is defined here as the effort expended by the human operator in accomplishing the task communication [35]. *In the field* refers to the assumption that humans are located in the same workspace as the robot, and can therefore use only the communication interfaces carried either by themselves or by the robot. *Conventional task communication methods*, against which we try to increase task communication performance, are reviewed in Chapter 2. The phrase *recognisable task sequences* refers to the assumption that most of the requested tasks

are part of task sequences that can be recognised, either from an *a priori* given or previously performed task sequence. However, task communication must also be usable should any number of unexpected tasks need to be requested.

The problem is limited to situations where only the two main task parameters, i.e. action and target object, need to be communicated. How certain other additional parameters might be communicated is not addressed. It can be argued, however, that it is much easier to define additional parameters after the main task parameters are known and the task context is thus defined.

Essentially, the task communication problem is also much simpler if the robot knows the exact sequence of tasks to be performed compared with a situation where the robot only knows, for example, the tasks it can perform. In the first case, the problem is merely to trigger the next task to be executed, when in the latter case, the task itself must also be communicated. The assumption here is that the robot has knowledge of certain possible task sequences, such as the task sequence required to set up a radio antenna, but it does not know when or whether the sequence will be performed. This knowledge about work context can be provided *a priori* or learned on-site while working. However, as was already stated earlier, any proposed approach must also be usable with completely unexpected task requests.

Another factor to be considered is the number of possible actions that the robot can perform with each object. In Chapter 3 of this thesis, five possible astronaut-robot planetary exploration missions are analysed, and the level of ambiguity in the action-object relationship and the number of objects are chosen based on those missions. In principle, almost any level of ambiguity and any number of objects could be considered applicable. For example, a robot capable of performing only one or two tasks, such as automated excavation, could be considered very useful in certain situations.

1.4 Research hypothesis and methodology

Although Natural Language Processing (NLP) has recently made good progress in some domains, such as IBM Watson did in the field of question answering [34], general all-context NLP is still a very challenging problem requiring complex human-like understanding of the situation. This means that we need to find new human-robot communication methods that are simpler, but still usable by humans. This thesis focuses on human-robot task communication, i.e. communication of at least the action and target object parameters of the task [90].

The hypothesis examined in the thesis is that humans are able to efficiently communicate tasks, consisting of actions and target objects [90], to a robot in the same way that humans can communicate tasks indirectly to other humans using only the task-related object or action names and by requiring the other human to associate the correct object to an action [136], or vice versa, i.e. action to object.

This hypothesis is tested with user experiments where participants are requested to communicate tasks to a robot first using only direct task requests, consisting of both action and target of action utterances, and then by also using indirect task requests, consisting of either action or target of action utterances. If the hypothesis is true, task communication times and the human workload should decrease when using an indirect task communication method instead of a direct one.

The human workload derived from task communication is measured using the National Aeronautics and Space Administration (NASA) Task Load Index (TLX) questionnaire [52], which rates the subjective workload from 0 to 100, i.e. from no workload to maximum workload. Task communication time is measured either as a time starting from when the task must be communicated to the beginning of communication, i.e. task request formulation time, or as a total time it takes to execute all the required tasks, i.e. total mission time.

1.5 Main contributions of the dissertation

The main contribution of this thesis is the novel method for human-robot task communication based on the concept of affordances. The idea of the affordance-based task communication method is to enable humans to use only a task's action or object names as indirect task communication requests and to let the robot perform the association of actions to objects, or vice versa, in order to define the entire task request.

More specifically, with regard to this proposed affordance-based task communication method, this thesis shows that

- It is technically feasible to implement fully autonomous human-robot interaction systems by utilising the proposed method.
- Humans are capable of communicating tasks to a robot using the affordance-based task communication method.
- Task communication workload and mission execution times can be decreased by using the proposed method.
- Humans prefer to utilise the proposed task communication method over current conventional task communication methods.

In addition, it was shown that task sequence prediction could be significantly improved by utilising partial communication in the form of only object or action names. This aspect of human-robot interaction has not yet been extensively researched.

Furthermore, a structured analysis of potential astronaut-robot planetary missions is presented. The novel analysis outputs are identification of potential astronaut and robot missions, and requirements for robotic astronaut assistants.

1.6 Author's contribution

The design and implementation of the affordance-based task communication method was carried out by the author. Likewise, the user experiments were designed and conducted by the author. Approximately half of the WorkPartner platform and software was developed for this thesis by the author and half by the General Intelligent Machines (GIM) research group.

1.7 Declaration of previous work

Parts of the work reported in this thesis have been published previously. The published material and publications, to which the author was the main contributor, are the following:

- definition of the most likely astronaut-robot mission scenarios and requirements for robotic astronaut assistants [56];
- review of human-inspired task communication methods [55];
- system architecture of the astronaut-robot cooperation system [54];
- user experiment results [55, 58, 57];

Other publications related to the thesis and the SpacePartner project had to do with the WorkPartner robot physics simulator [59] and the manipulator algorithms for physical human-robot interaction [140]. However, these publications deal with topics that are not directly addressed in this thesis.

1.8 Thesis outline

Chapter 2 presents a review of related work, starting from the relevant terminology and previous research done with robotic assistants testbeds. Next, the possible user interfaces for communicating with robotic assistants are reviewed, along with human-inspired Human-Robot Interaction (HRI) methods. The review finishes by presenting in detail the concept of affordances and how it has been applied in user interface and robotics research.

Chapter 3 presents analysis focusing on identification of the most likely planetary exploration missions involving robotic astronaut assistants, and the requirements these missions pose for these robots. Finally, these robot requirements are compared with the WorkPartner robot in order to understand robot readiness to perform the identified tasks.

Chapter 4 introduces the field experiments done with the actual fully autonomous WorkPartner robot. In these experiments, the work environment is restricted so that each object is unambiguously associated with only one action, and vice versa.

Chapter 5 builds on the experiments in Chapter 4 by utilising a simulated WorkPartner robot in more complex work environments. These experiments extend the work environment to situations where the object-action associations are ambiguous.

Finally, Chapter 6 concludes the thesis, and Chapter 7 presents ideas for future work.

2 State of the Art Regarding Robotic Assistants

This chapter presents a review of previous research related to robotic assistants. Section 2.1 and Section 2.2 start by introducing the relevant terminology and previously developed robotic assistants, respectively. Next, possible user interfaces for communicating with robotic assistants are reviewed in Section 2.3, while human-inspired HRI methods are examined in Section 2.4. Section 2.5 presents in detail the concept of affordances, and how it has been applied in user interface and robotics research. The chapter concludes with remarks in Section 2.6.

2.1 Terminology

It is important for any scientific document to be as clear and as unambiguous as possible with respect to the terms used in the report. Writing exact term definitions is, however, especially important here because the field of HRI research is relatively young and interdisciplinary. In most cases, there are no exact general de-facto definitions of terms, which is why we can consider the given definitions as “working definitions”, i.e. definitions chosen to allow the work to proceed even though it is understood that they are not complete or final. The purpose is to have term definitions that are useful in the context of this thesis.

2.1.1 Action, task and mission

One of the key terms in this thesis is task. Task has been formally defined as “a set of (human) actions that contributes to a specific functional objective and ultimately to the output goal of a system” [115]. A task can usually be decomposed into more elementary subtasks while a set of tasks forms instead a higher-level mission

[33]. Missions, tasks, and subtasks are all different levels of activities, as shown in Figure 2.1. In this thesis, task is defined as consisting of task-related action and a target object, which can be considered the minimum of parameters needed to define a proper task [90]. Action has been defined in literature, for instance, as “move from one state to another, in order to achieve the desired state” [16] or as “doing something to the world - move yourself or manipulate someone or something” [96].

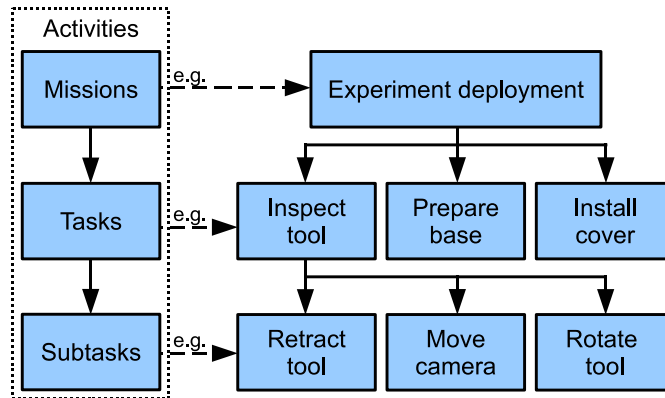


Figure 2.1: Missions, tasks, and subtasks are all different classes of activities.

Mission is the highest activity level and is constructed based on mission objectives [33]. Examples of missions are: science experiment servicing and rover sample acquisition. Mission objectives are also referred to as mission goals [33]. “Goal” itself can simply be specified as “something that we want to achieve” [96].

As stated at the beginning of the chapter, the definitions of terms presented are not consistent in the literature. For example, a broader definition for activities could be “located behaviours, taking time, conceived as socially meaningful, and usually involving interaction with tools and the environment” [16]. The above definition is given in the context of modelling group behaviour, and although such definitions could be selected, it can be argued that the added complexity of the terms would not add value in the context of human-robot task communication.

2.1.2 Collaboration, cooperation and coordination

Collaboration, cooperation, and coordination are related terms that are used to describe how robots and humans perform activities together. Collaboration has been defined, for example, as a process where two or more robotic or human actors work together to achieve shared goals [35]. On the other hand, it has also been argued that collaboration requires consciousness, meaning that collaboration would be - with current robotic technologies - restricted to humans [17]. This thesis adopts the collaboration definition presented in [27], which states that collaboration is “coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem”.

Cooperation is considered to differ from collaboration by the way the work is divided [27]. In cooperation, the work is divided into independently solvable sub-tasks and coordination between actors is needed only to combine the results. Collaborating actors work together simultaneously, rather than independently, and they also use coordination to define and divide the work. For example, negotiation and argumentation are types of interaction typical of collaboration.

Coordination is an important part of both cooperation and collaboration. Coordination can simply be defined as managing dependencies between activities [81]. If there are no dependencies between actors, then there is no coordination, because there is simply nothing to coordinate. Dependencies can, for instance, be shared resources, producer/consumer relationships, simultaneous constraints, or tasks. Correspondingly, coordination processes that can manage these dependencies are, for example, priority ordering, sequencing, scheduling, and task decomposition.

From the above definitions it can be summarised that cooperation has primarily task execution-related dependencies that need to be managed, e.g. timing and solving anomalies, and collaboration typically has task definition dependencies that

need to be managed, e.g. refining goals and allocating tasks. This means that collaborative actors characteristically elaborate the shared work as they proceed, while cooperative actors focus on properly executing or operating the defined joint work.

The various described levels of dependency management in coordination, cooperation and collaboration are shown in Figure 2.2. In this thesis, the astronaut and the robot cooperate to perform activities, which consist hierarchically of missions, tasks, and subtasks.

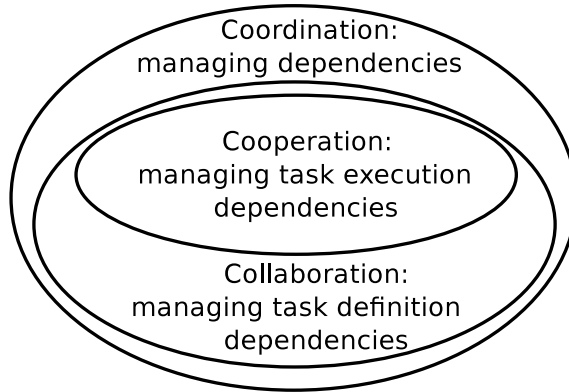


Figure 2.2: Coordination, cooperation and collaboration are all managing dependencies.

2.1.3 Interface, interaction and communication

For coordination, cooperation or collaboration to be possible, humans and robots need interfaces in order to interact. Interface is defined in a telecommunication glossary as “a shared boundary, i.e., the boundary between two subsystems or two devices” [131]. An interface can thus be seen as a boundary that enables separate systems to connect and interact with each other.

By comparison, interaction has been defined, for example, as “mutual or reciprocal action or influence” [86]. One common aspect of interaction definitions is the focus on the relationship of two or more entities [112]. For example, based on the above definition, HRI is the mutual action or influence of human and robot entities enabled through a human-robot interface. This definition of HRI is used in this thesis. A similar but slightly more abstract definition of HRI is related to the study of how humans and robots influence each other [46].

Communication is another term that is related to interaction and interfaces. Communication is defined, for example, as a process of information exchange between actors [130]. Communication can be seen as the “mutual influence” part of the presented definition of interaction. This means that human-robot communication is the part of HRI that deals with information exchange. Human-robot task communication, which is one key term of this thesis, is thus the exchange of task-related information between human and robot.

2.1.4 Natural and intuitive human-robot interaction

Another term used in this thesis that lacks a proper, commonly accepted definition is natural HRI. Natural interaction has been linked, for instance, to ease of learning [2] and to reducing fatigue and sickness in simulators [116]. In simulated environments, natural is used to refer to a resemblance to real world interaction in the context of use, e.g. a tennis racket is a natural way to interact with a tennis game [100].

However, the most common definition is probably the one that holds that natural interaction refers to interaction that we can observe between humans [105, 50, 63]. Natural interaction thus utilises human modalities such as gesture, speech, touch, vision and smell [87]. This thesis adopts this definition and uses natural interaction as a synonym for human-to-human-like interaction.

Based on the above definition, the goal of natural HRI is seen to be strongly linked to identifying human communication characteristics and applying them to improving communication [2]. We need to understand how humans communicate and process information in order to be able to build robots that are compatible with human communication.

Another term very similar to natural is intuitive. Intuitive interaction has been used, for example, as a term referring to efficient interaction without conscious use of previous knowledge [94]. This requires the interaction to be easy to learn and to remember, both of which are also features of natural interaction. The terms intuitive and natural are differentiated in this thesis based on the idea that natural interaction must be inspired by interaction between humans. Thus, for example, a steering wheel is an intuitive way to interact with a car, but it is not natural.

2.2 Robotic assistants

At present, the only operational robotic astronaut assistants are the remote manipulators used by the space shuttle and the International Space Station (ISS). These tele-operated robots are used as crane-like manipulators to transfer Extra-Vehicular Activity (EVA) astronauts and payloads [101]. Human space exploration over the next decades will focus on the surfaces of the Moon and Mars. This means that new types of astronaut assistants are required, especially on Mars where tele-operation from Earth is not viable due to the long communication delay. The purpose of this section is to present the current status of the development of robotic EVA astronaut assistants, focusing on the surfaces of the Moon or Mars.

First, it would be a good idea to define what the term robotic astronaut assistant actually means. The term assistant is defined as “a person who contributes to the fulfilment of a need or furtherance of an effort or purpose” [106]. On the basis of

this definition, a robotic astronaut assistant is a robotic actor that contributes to the fulfilment of an astronaut's effort. This relatively loose definition is enough to trigger important follow-up questions: what advantages could assistance offer; what kinds of robots could be used for assistance; and what are the efforts or activities to be assisted?

The potential of robotic astronaut assistants has already been recognised. For example, both the NASA and the ESA have identified crucial roles for various kinds of automated and robotic technologies in their future space exploration missions [122, 110]. The overall motivation for providing robotic technologies for crew assistance is to extend the crew's capabilities during exploration missions [110]. This extended capability can be seen in terms of a combination of increased scientific output and crew safety, as well as a decrease in the overall mission cost and crew workload [114, 110, 24].

Many different types of robotic astronaut assistants are considered suitable for space exploration missions. For example, it has been stated that both micro (1 to 20 kg) and mini (20 to 150 kg) rovers are essential for robotic and human planetary exploration [110]. Another view is that humanoid robots are "key partners" to be considered for construction and maintenance because of their form [123], which enables them to perform in environments designed for humans. It has also been argued that a wheeled centaur-type robot configuration is desirable in order to guarantee both dexterous manipulation capabilities and mobility on rough planetary surfaces [85].

In the end, the right mass, shape, strength and flexibility for a robotic assistant depends on the activity to be performed [110, 123]. There is, however, general acceptance that tasks such as construction, assembly, and maintenance would require the robot to have at least some level of intelligence, autonomy, mobility, depth vision, and manipulation ability [123].

The robotic assistant's level of autonomy is what ultimately determines how the tasks can be divided between astronauts and robotic assistants. In the ideal case, robots could take care of all tasks if necessary, while in the worst case, robots are not able to perform any useful tasks. For example, it has been stated that the level of dexterity of an EVA astronaut will be achievable by tele-operated robots in the near future, but not by autonomous robots [21], while automated inspections, on the other hand, could be viable in the near future. The biggest challenges to be met in autonomous robotic operations are robustness in complex environments and human-level adaptability [21].

2.2.1 Research in astronaut-robot cooperation

First steps in robotic astronaut assistant development were taken by the NASA Ames Research Center (ARC)'s Astronaut-Rover (ASRO) project in 1999 [14, 12]. The project target was to identify activities where humans and robots could work as a complementary and interactive team, and identify the requirements for such cooperative rovers in order to support safe, productive, and cost-effective surface reference mission development.

Using a tele-operated Marsokhod rover, as shown in Figure 2.3, the ASRO project performed four missions representing potential astronaut-robot interaction missions. These missions were: pre-EVA scouting; video documentation; field science experiments; and assistance in transporting objects. The missions tested identified needs for enhanced astronaut-robot communication, for example, using voice, visual observations and target marking beacons, and the requirement that the robot be at least as fast as the astronaut.

The ASRO project's research in astronaut-robot interaction and cooperation was continued by the NASA Johnson Space Center (JSC)'s EVA Robotic Assistant



Figure 2.3: The astronaut and the Marsokhod robot cooperated in the ASRO project.

(ERA) project [12]. The ERA project was aimed at producing a robot that can assist spacesuited humans and to provide design constraints for Mars reference missions. Three representative surface exploration activities requiring astronaut-robot cooperation were tested in year 2000 using a modified iRobot ATRV-Jr mobile platform, shown in Figure 2.4, which operated mostly autonomously. These activities were power cable deployment, solar panel deployment, and object transportation. The activities tested identified the following robot core capability development needs: communication, manipulation, and navigation.

Several other field tests were also performed during the ERA project [13, 18, 43]. The ERA rover, named Bordeaux, used in the tests was enhanced with a Metrica Inc. robotic arm and with a three-fingered hand. The activities performed in the field tests were geophone¹ deployment, astronaut tracking and monitoring, location tracking and place naming, science data logging with photos and voice, biosensor logging, activity duration and sequence tracking, picture-taking on command, reconnaissance (solo-scouting), communication network relaying, and providing a remote workstation for astronauts. The main findings were the need for force sensing on

¹Geophone is a device which converts ground movement into electrical voltage.



Figure 2.4: The ERA project’s robotic astronaut assistant.

the manipulator and the need to strongly integrate several systems by, for example, using multi-agent architectures.

The most sophisticated robotic assistant ever developed is probably NASA’s human-size Robonaut, shown in Figure 2.5. It is a wheeled humanoid robot with more than 40 Degree Of Freedom (DOF) and designed to achieve a spacesuited astronaut’s dexterity [4, 26]. The goal of Robonaut development is increased astronaut safety [26], and ultimately, the ability to provide a human cognitive presence without human physical presence [21].

A wide range of different activities has been tested with the Robonaut in teleoperation mode. These activities include cable deployment, rock sample collection, metal beam alignment, tying a knot, and locking an electrical connector. Tests performed indicated, among other things, the need for compliance control in manipulation and the need to intelligently divide the work between robot and astronaut.



Figure 2.5: NASA's Robonaut picking up objects and attaching tether hooks.

Autonomous Robonaut capabilities were later developed based on the ERA software [25]. Automated activities tested include navigation, acquiring tools from humans, and following humans. The main test results indicated a need for robust communication channels and capabilities to deal with confusing sensor data.

NASA's probably most advanced human-robot EVA operation test scenario involved two suited astronauts and four robots [24]. The test scenario was set up to assess the types of tasks robots can perform on the Moon or Mars surface environment. The performed robot-astronaut scenario tests included: autonomous robotic payload removal, stowage operations under local and remote control, and autonomous robotic navigation and inspections [24]. The NASA JSC's Centaur robot was successfully used in supervised autonomy mode for the astronauts' rover unloading task, and the NASA ARC's K-10 robot was used for visual inspection of the rover. Other robotic tasks tested were hill climbing, moving heavy loads, gathering geological samples, drilling, and tether operations. The test results explicitly demonstrate the feasibility of human-robot team cooperation for EVA surface exploration activities.

The ESA has also researched planetary astronaut-robot cooperation with the so-called Eurobot Ground Prototype (EGP) platform [141], which is shown in Fig-

ure 2.6. Their research goal was to analyse the feasibility of human-interactive on-surface operations of a centaur-type robotic system with extensive manipulation capability. Thus, tasks studied included, for example, handover of items, and transportation of heavy equipment. The experiment’s most important observation related to human-robot interaction was that vision-based environment perception is still a serious challenge, which could, however, be partially solved by requiring astronauts to communicate and point out target objects to the robot.

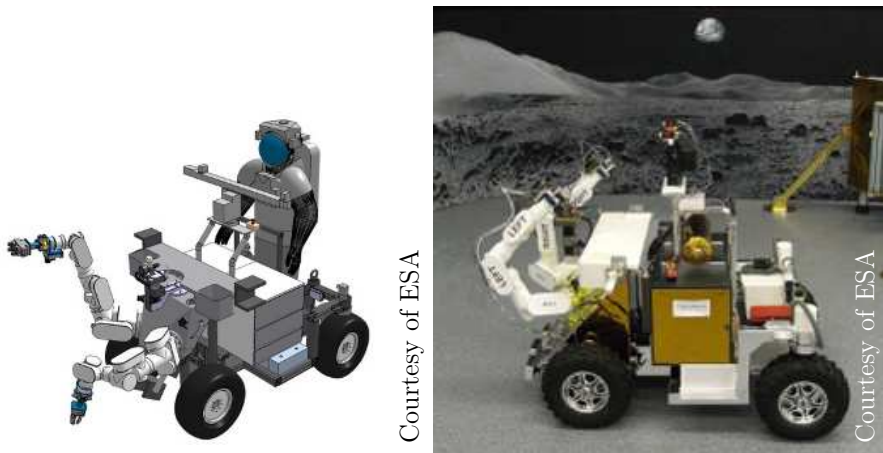


Figure 2.6: Eurobot Ground Prototype, an ESA platform for analysing the feasibility of human-interactive on-surface operations.

2.2.2 Research in human-robot cooperation

Human-robot cooperation has also been actively researched for non-space applications. This section describes some of the experiments where non-astronaut humans cooperated with robots.

One idea that has been introduced to enable cooperation between humans and robots is the so-called dialogue-based user interface and control architecture [35]. The moti-

vation underlying this work was the use of human and robot capabilities when most appropriate, i.e. robots are good at structured decision-making and repetitive work, while humans are better at unstructured decision-making, object recognition and situation assessment. The work presents a PDA-based GUI that enables robots to use humans as a resource by asking questions while executing a remote driving task. The resulting system is considered to coordinate robot action, facilitate adjustable autonomy and human-robot interaction, and to enable the humans to compensate for inadequate robot autonomy. This kind of peer-to-peer human-robot interaction is presented in more detail in Section 2.4.3.

The idea of using the most suitable robotic and human capabilities when appropriate has also been examined in a heterogeneous robot cooperation framework [118, 119]. The development motivation is to enable the use of robotic labour where the use of human labour is hazardous, expensive or scarce. This work on sliding autonomy tries to solve the following three problems: when to call a human for help; how to provide situational awareness to the user; and how to maintain work coordination after human intervention. The problem of when to call for help is approached by using performance models of actors, information about human learning curves, and information about the team state.

Heterogeneous robot cooperation was tested by means of experiments where a square structure was assembled using four beams [118, 119]. The first experiment examined four coordination strategies: pure autonomy, System-Initiative Sliding-Autonomy (SISA), Mixed-Initiative Sliding Autonomy (MISA) and tele-operation. The results of the experiment showed that autonomy was faster but less reliable than tele-operation. MISA and SISA improve system performance by increasing reliability while still being almost as fast as autonomy. The second experiment examined the amount and type of information required to minimise the time needed to achieve situational awareness. The performance achieved was shown to be a trade-off between time used and quality of understanding.

2.3 Human-robot interfaces

The purpose of this section is to review developments in human-robot interfaces. Available interfaces set practical constraints on which type of communication methods can be implemented. For instance, the availability of accurate mind-reading interface devices would probably change how we communicate with robots and computers [126].

The section is structured according to the interface types, namely: gesture, visual displays, speech, and haptics. Each interface type is examined with regard to its suitability for conveying different types of information, especially in the examined astronaut-robot planetary exploration application.

2.3.1 Gesture interfaces

One way to convey information is to capture human body movements and to utilise them for human-robot communication. Existing motion capture technologies - also referred to as motion tracking technologies - can be divided into six categories plus a hybrid [138], as shown in Table 2.1. Of interest from a space exploration perspective are the portable ones: image-based, inertial, and mechanical. In particular, an image-based approach would probably be used anyway because robots and humans will very likely have cameras with them. For this reason, the rest of this section deals with image-based approaches. The methods for transforming captured motions for different use applications are nevertheless applicable to all of the motion capture technologies.

Most of image-based human motion analysis is divided into human detection, tracking, and behaviour recognition [139]. The goal of human detection is to separate humans from the rest of the image through motion segmentation and object clas-

Table 2.1: Comparison of motion capture technologies. The positioning accuracy varies from one millimetre (optical) to a few centimetres (acoustic).

Technology	Principle	Example	Portable	Positioning
Optical	Reflective or emitting markers	Vicon	Yes/No	Absolute
Magnetic	Pose in magnetic field	MotionStar	No	Absolute
Image-based	Tracking of image features	EyeToy	Yes	Absolute
Inertial	Inertial Measurement Units (IMUs)	Moven	Yes	Relative
Mechanical	Joint-angle measurement	ShapeWrap	Yes/No	Relative
Acoustic	Audio signal Time Of Flight (TOF)	Bat	No	Absolute
Hybrid	Combine several methods	Hy-BIRD	Yes/No	Both

sification. The motion segmentation task detects moving segments in the image by using, for example, background subtraction, and the object classification task classifies objects based on, for example, object shapes.

The objective of human tracking is to find relationships between objects among consecutive image frames [139]. Tracking approaches can be divided into model-based, region-based, active contour based, and feature-based approaches. Model-based approaches represent the human body as a number of geometric objects connected with joints. Region-based approaches identify regions and use cross-correlation to track the region between images. Active contour-based tracking is similar to the region-based approach, but extracts the shape of the target tracked between images. Finally, feature-based approaches do not try to track objects, but track instead single features like points or lines. Popular mathematical tools for human-based tracking are Kalman filters, Hidden Markov Models (HMMs), cross-correlation calculations, and particle filters.

The goal of human behaviour recognition is to be able to classify human actions based on tracked human motions [139]. Behaviour recognition can be done, for example, by matching tracked data sets to action templates or by defining paths

in state-space models, e.g. in HMM, as specific actions. Behaviour recognition is a common problem for all motion capture technologies.

Bayesian networks, for example, have been used to provide a solution to human action classification [117]. The idea in this case was to create an intention recognition model using intention-action mapping where a human expert is utilised to associate actions with intentions. The Bayesian network gives estimates for the likelihood of different intentions when certain actions are observed. The system was tested in a virtual kitchen environment, and it was shown to be capable of recognising user intentions.

In addition to understanding astronauts' gestures, robots can also execute attention-drawing gestures. For example, a five-step natural communication process was developed to enable attention-drawing between human and robot [127]. Presented communication processes are context focus establishment ("we talk about boxes"), attention synchronisation ("robot tries to look in the direction indicated"), object recognition ("robot looks at the object and makes a sound"), believability establishment ("human corrects if something went wrong"), and object indication ("robot points to the object"). The five-step process was tested in an experiment where a human communicates object locations with speech and gestures. The results showed that the five processes increased recognition of the objects indicated.

A vision-based human motion analysis can be also done using a wearable camera [51]. Head-mounted cameras are capable of tracking human hand trajectories and can thus be used, for example to recognise object manipulation activities or hand gestures. These cameras worn by humans can also be used for other purposes like activity documentation or providing Situation Awareness (SA) to other actors. In addition, human-mounted cameras could be used as input devices in GUI. These kinds of wearable cameras could also be readily embedded in astronaut spacesuits.

2.3.2 Visual displays

Visual displays are considered here as a wide range of different ways to produce visual information, probably best described as a combination of Virtual Reality (VR) and Mixed Reality (MR) [88]. MR refers to both Augmented Reality (AR), i.e. the real environment is augmented with virtual objects, and Augmented Virtuality (AV), i.e. the virtual environment augmented with real objects. The mix ratio of virtual and real environments can be described, as shown in Figure 2.7, using the so-called virtuality continuum.



Figure 2.7: The virtuality continuum describes the mix ratio between real and virtual environments, adapted from Milgram et al. [88].

One interesting form of AR is to overlay information on a camera image. For example, Rosenstein et al. [111] present a method that interprets operator intentions based on control of the robot in the vicinity of landmarks. This is done using virtual geometric objects, called funnels, which provide artificial landmarks for the operator. The operator can, for example, change the level of autonomy only when the correct landmark is activated, and operator intention is thus recognised.

It is also possible to use visual input directly in the environment by, for example, using a laser pointer to select objects from the real world [71]. In this case, the laser point is the visual artefact that can be displayed and recognised by both robots and humans, i.e. it acts simultaneously as input and output device [137]. The laser point can be recognised using stereo cameras with an average accuracy of around ten centimetres when operating from a distance of three meters [71].

2.3.3 Speech interfaces

Speech is probably the most common way of exchanging information between humans. It is also very probable that an astronaut always has a microphone available, so use of speech interfaces is a very likely choice. Speech can be used to give information about a state, especially warnings, and to give names and descriptions of objects.

Voice-controlled systems have already been tested on simulated Mars exploration missions [19, 121]. For example, predefined spoken commands were used to communicate with a software agent system whose main purpose was to provide information about ongoing events [19]. The performed tests did not examine what would be the best way to give different task requests, but argued that task requesting would be developed further later. Their tests indicated, however, that voice would be a good way to receive many different types of information on planetary space exploration missions. Numbers were identified as one exception that needed to be supported with, for example, a watch-like device.

It is also possible to use previously unknown commands and words, such as names, with a robot. For example, Funakoshi et al. [41] present a location-naming system using speech interaction. The core research problem, called out-of-vocabulary problem, is that the places named are words that are previously unknown to the robot. The problem then becomes how to assimilate the meaning of new words. The solution presented is word classification using the Bag of Words in a Graph (BWG) method. Basically, the location names are saved as audio signal frequency patterns, and similarity is used for recognition.

Name teaching is done in special “learning mode”, while robot commanding is done in so-called “execution mode”. The reported average word recognition rate of the system is 83.3%. This kind of out-of-vocabulary naming of objects could also be

necessary for the work presented in this thesis, when humans and robots are working with previously unknown objects.

2.3.4 Haptic interfaces

The term “haptic” is used here to refer to the sense of touch, position, motion, and force [39]. It is a broader term than a related “tactile” term that refers just to the sense of touch. For example, a vibrating cell phone or buttons attached to a robot can be considered haptic interfaces. They can be used to communicate, for example, motion trajectories, activity progress, or changes in the level of autonomy. Some haptic interfaces might be usable only when actors are in close proximity with each other, such as buttons on robots, while other haptic interfaces can be used or activated over distance, such as vibrating cellphones.

One of the simplest haptic interfaces is where a human gives force and motion as input and the robotic system replies with counter force. For example, Hirata et al. [60] present a walking assistant which interacts with humans by using two servo brakes on the wheels. Using the servo brakes, the walking assistant can steer the movement which is created by the user. It is considered to be a safe way to interact, because all the system energy is created by the user.

Haptic interfaces can be useful in extreme situations where other communication interfaces are not practical anymore. Naghsh et al. [92] performed a robot swarm assisted fire-fighting scenario with both haptic input and output interfaces. A tactile interface, which was used as a human output device, used eight tactors on the fire-fighter’s torso to communicate possible hazards through frequency and amplitude signals. Large buttons were mounted on top of the robots to act as robot input devices. The buttons were used to control otherwise autonomous robot swarm activities.

2.4 Human inspired human-robot interaction

It has been reported that 71% of people would like robots to be able to communicate in a more human-like manner, but only 36% and 29% of people would like the robot to behave and appear more human-like, respectively [22]. This means that the key issue is natural human-robot communication, rather than making the assisting robot itself more human-like.

After a short introduction to HRI design, this section presents and analyses four different approaches that have been taken to develop natural human-robot interfaces. The common element in these approaches is the way user interfaces have been inspired by task communication between humans. There are many other human-robot interaction aspects, such as social robots, that are not examined here because there is in practise very little experimental evidence available in favour of their usefulness in astronaut-robot cooperation contexts [37].

For each of the methods presented, a system-level user interface module diagram is shown in order to clarify exactly what the presented methods do in practise. A general view of such module diagrams is shown in Figure 2.8. Essentially, the user interface module deals with communication with humans, and transforms information into a format suitable for the robot.

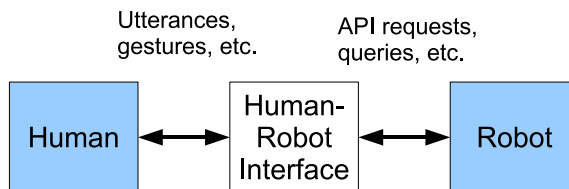


Figure 2.8: A general user interface module diagram.

2.4.1 Design of human-robot interaction systems

Some special considerations are required in robot system design when humans are introduced as a fundamental part of the system. These needs have emerged as traditional industrial robots have evolved into service robots that share physical space with people [133]. The new aspect of a strong human presence and involvement provides one of the key roles in HRI system design.

As with traditional robot systems, frameworks must be designed to support HRI development [32]. Their purpose is to provide basic services such as data transfer, support different display views, and facilitate human-centred interaction. The development of design paradigms for a human-robot system is still ongoing, because although there have been efforts to develop such systems, none of them have actually met the requirements set [32].

It has also been stated that effective, efficient and natural HRI is crucial to the success of future space exploration missions [32]. In this case, human-robot system design challenges are mostly related to information exchange. It is considered that robots and humans require the ability to communicate about their goals, abilities, plans, and achievements. Robots are also expected to interact with humans, both locally and remotely, to solve problems that exceed their autonomous capabilities.

One of the current approaches to human-robot design is to emphasise the importance of designing human-robot interfaces to meet human needs [1, 91], i.e. to take into account ergonomic issues. These ergonomic issues include support for human decision-making processes; achieving proper workload levels; maintenance of situation awareness; and minimising the possibility of human error.

2.4.2 Natural and indirect human-robot communication

There have been several attempts to enable humans to communicate instructions to robots by using natural language [28, 5]. These systems all essentially process utterances according to certain syntactic and semantic rules into robot-usable commands and information, and are as such still far from an unrestricted human-human type of communication [142].

As a consequence of this requirement to use a rather fixed set of possible utterances, there will be cases when human utterances are not understood by the robot, either partially or completely. The most common approach in this case is to initiate a dialogue in which the human can define the missing information using either a proposal list generated by the robot [66, 77], or by answering an open-ended question [142].

Consequently, task requests selected from a robot-generated list with an explicit natural language, such as “analyse rock”, are considered in this thesis to be conventional task communication methods, against which the proposed affordance-based task communication methods are being evaluated.

Some level of indirect task communication has already been used for human-robot communication in the form of completing empty task parameters with default values [10]. In general, a task is essentially communicated indirectly every time a robot automatically completes some part of the task request.

The difference with task execution ambiguity solving, such as how to decide which way to go past an obstacle when it was not explicitly specified, is very hard to define exactly. It is a question of defining the level of detail required for task communication. In this thesis the requirement was to communicate the task-related action and target object parameters.

Automatic ambiguity solving in task communication has been tested, for instance, based on the object’s spatial distance from the human actor [77]. This means that if there are several target objects that could be referred to, then the closest one will be the one selected. However, this is more a kind of execution ambiguity solving, as both the action and target object parameters are being communicated.

One other option for resolving ambiguity is to utilise previous task requests [142, 10], sometimes combined with information about a robot’s past movements [15]. For example, “more” would mean “more forward” if the previous task was to move forward. However, this is in principle further communication of a previous task request rather than communication of a whole new task request, because the human needs to eventually communicate all the task parameters.

The concept of context predicates is very similar to the previous one where a stack of actions is used to complete partial task requests [103]. The idea of context predicates is to enable interruption and continuation of execution toward several different goals, instead of just one previous goal, by preserving past tasks in a stack until they have been successfully completed. This approach could be also applied to enable several tasks to be requested for execution with the indirect task communication method presented in this thesis.

2.4.3 Peer-to-peer dialogue

Humans providing assistance rarely only take and perform requests; they also actively observe the situation and consider the appropriateness of the communication. If something is not as it is supposed to be, humans do not just stop, but instead decide if a new dialogue should be initiated to solve the problem. This type of communication dialogue is referred to as peer-to-peer dialogue [36, 69], meaning that communicating actors are considered equal as they are both able to initiate dialogues.

This type of human-human-inspired dialogue has been developed to interact with robotic actors [35, 38]. The dialogue system idea is intended to enable robots to be human peers, as they can use the humans as a resource by asking questions while executing tasks, just as people do. In this way, dialogue can be seen to enable the use of both human and robot capabilities when they are most appropriate. For example, robots are good at structured decision-making and repetitive work, while humans are better at unstructured decision-making, object recognition and situation assessment.

The peer-to-peer dialogue idea was first tested in an office environment with a tele-operated exploration robot [35]. Evaluation of the test indicated that dialogue was especially helpful in allowing humans to understand the problems the robot tried to solve. However, at least when the human actor is in the front of a tele-operation station and focused on the robot, it did not seem to be very necessary for the human to ask the robot questions, as all the information was already available. Requesting the robot perform tasks was, of course, a very important part of the dialogue.

It can be argued that the main advantage of a peer-to-peer dialogue system is the sharing of the actors' knowledge and capabilities. The robots can utilise the superior human cognitive capabilities and the humans can incrementally communicate task parameters to the robot if needed.

Some of the disadvantages of the dialogue have been found to be too frequently asked questions and possibly irrelevant questions [35]. This indicates that the robot's threshold for asking the human actor questions should be adjustable. In the end, dialogue system performance is very much dependent on the robot's ability to evaluate whether the human actor should be addressed or not.

The peer-to-peer system could also be incorporated as a user interface module, as shown in Figure 2.9. In a case where the robot is requested to perform a task,

the peer-to-peer user interface module can check if all the parameters required by the robot were given, and if not, they can be requested through the dialogue. If, instead, the robot needs help, the peer-to-peer user interface module can initiate the dialogue with the human should the importance of the event exceed the currently used threshold level.

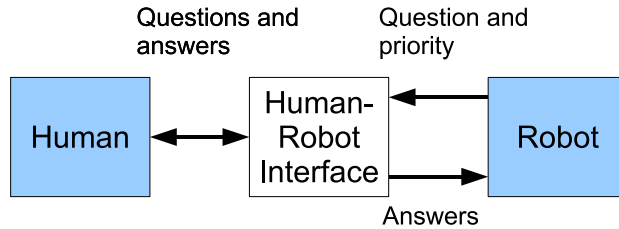


Figure 2.9: The peer-to-peer dialogue communication method described as a user interface module.

2.4.4 Perspective taking

One distinctive feature of human-human communication is the description of spatial locations relative to other actors or objects. For example, objects can be described to be on top of other objects or on a certain side of the questioner.

The exact coordinates, which are usually required by the robot, are in practise never used. In fact, according to the analysis of two astronauts training for an ISS mission, 25% of the time astronauts had to take the perspective of other astronauts into consideration [38].

Human-robot interaction based on perspective-taking has been researched and tested in a few different applications [38, 134, 120]. Nevertheless, the basic idea is the same: perspective-taking enables the robot to reason and simulate the world from the per-

spective of others. Using perspective-taking, the robot can limit the possible action options that the user could refer to during task communication based, for instance, on the objects' visibility to the user.

The ability to understand perspectives has been implemented and tested in complex real-world experiments [134]. In 20 different trial runs, the robot was shown to be able to simulate object visibilities from the perspectives of others and by using different actor and object reference frames to make correct decisions about, for instance, exactly which cone the person might be referring to.

The main advantage of the method is added flexibility in describing spatial locations. Although the descriptions can be quite broad, in most cases they are sufficient to limit the options to one unambiguously defined object.

The disadvantage of the system is that the robot has to maintain a relatively accurate model of the environment with moving actors and objects. This can be a computationally heavy task, especially if the environment is complex. Some type of environment model would be needed in any case, even if the target were to be described from the robot's perspective, so the disadvantage is quite marginal.

A user interface module incorporating perspective-taking requires only information about object and actor locations in order to be able to tell the robot exactly what needs to be done. Such an environment map interface is often readily available, so the perspective-taking functionality could be included to a user interface module. A system-level view of how perspective-taking could be built-in as a user interface module is shown in Figure 2.10.

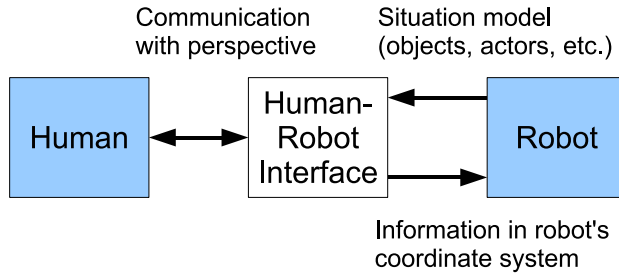


Figure 2.10: The perspective-taking communication method described as a user interface module.

2.4.5 Common ground

Shared knowledge and beliefs have long been identified as fundamental requirements for successful communication between humans [68, 20]. This assumed shared information is referred to as common ground, and the process of establishing it is referred to as grounding. For example, an utterance such as “this place is the goal” can be used to create shared knowledge about a location, called goal, which can then be used in subsequent communication.

Different approaches to establishing common ground have been incorporated into robots to improve task communication [79, 124, 125]. The goal of this incorporation is to have a set of shared information that can then be used to communicate tasks. With task-relevant common ground, the amount of communication is minimal while still being unambiguous about the task to be performed.

For example, a robot has been designed to learn basic concepts in a private house, such as kitchen and a favourite cup [79]. These mutually understood concepts were then successfully utilised to ask the robot to perform tasks. This grounding could also be done during task communication, instead of being performed in advance.

Another design introduced a robot that builds common ground between the user and the robot by asking further questions to clarify the given task plans in case they were potentially ambiguous [125]. The tests performed showed that such a grounding process helped to decrease the number of erroneous task plans given to the robot.

The main advantage of common ground is the decreased amount of task communication required. This is because having many mutually understood concepts permits less detailed task requests to be used [72]. If it cannot be assumed that the robot knows, for example, the names of the rooms, then communication has to rely on more general and abstract concepts, which increases the communication effort.

One difficult issue in establishing common ground is to know how to spend just the correct amount of time to ground all the required shared information. If a task needs to be communicated only once, then it does not make sense to use too much effort to establish common ground. For example, people do not start to teach a tourist about places in the city when explaining directions, but use only more general descriptions. The challenge here is to find the optimal balance between the effort and the time used for grounding and for task communication.

What is required of a robot module that can do grounding and utilise common ground in task communication? In the end, it requires only that a model of the current situation be available. The common ground module can update this situation model, for example, by naming certain locations, and use the model to transform task communication into a format that the robot can directly utilise. The peer-to-peer communication described in Section 2.4.3 could be used, for example, to acquire any missing information that still needs to be grounded. A system-level view of how common ground could be build as a user interface module is shown in Figure 2.11.

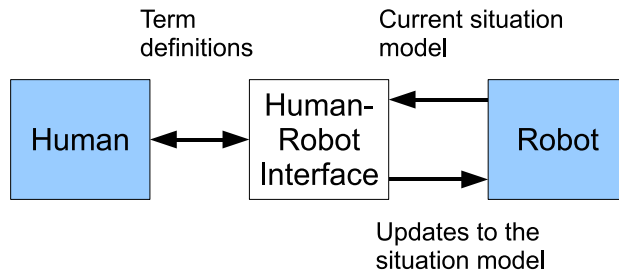


Figure 2.11: Establishing a common ground described as a user interface module.

2.4.6 Deictic terms and gestures

Terms, such as “that” or “there”, whose meanings depend on the current situation, are also part of human-human communication [78]. These so-called deictic terms or references are ambiguous by themselves, so complementary information must be given [82]. This complementary information can, for instance, be given through deictic gestures [82] such as a gaze or finger pointing, or obtained through analysis of the situation [78, 75], such as previous utterances.

The frequency of use of deictic terms and references in human-human communication has been found to be important in certain types of situations. For example, in a situation where speech utterances and pointing gestures were allowed when explaining the wiring of network equipment, pointing gestures were used over 90% of the time, and in about 40% of these instances they were also accompanied by deictic speech utterances [6]. Another test showed that over 50% of spontaneous hand movements were deictic gestures, i.e. pointing towards objects or actors, in a situation where a person described a painting to another person without any visual contact between them [48].

Deictic references have been the topic of research both in human-computer interaction [8] and in human-robot interaction [11]. The overall goal is to complement ambiguous deictic task communication terms with deictic gestures. Deictic gestures have been shown to be preferred over speech descriptions for certain tasks, such as for guiding workers through physical tasks [6].

Most of human-robot research with deictic communication has focused on communicating tasks to the robot. For example, a wheel-attaching task was tested in an astronaut-robot interaction context, showing that functional implementation using deictic referencing can be done and is usable [11]. However, deictic gestures have also been incorporated to enable the robotic actor to make gestures [128]. In this case, a robot was shown to be capable of pointing out targets to a human actor, in addition to using verbal communication.

The main advantage of deictic gestures is that they provide a mechanism to make task communication unambiguous, most typically speech [11]. Without deictic gestures, the ambiguous deictic references in communication would have to be replaced, for example, with verbal spatial descriptions. Deictic terms such as “this”, provide a way to directly link the deictic gestures to other communication.

Nevertheless, it is no small matter to accurately extract the deictic gestures without incorporating relatively complex sensor mechanisms [102]. In practise, this added system complexity means vulnerability, which indicates that deictic references should not be the only communication method, but rather an additional method. Pointing gestures are also not usually very exact, so they might be only sufficient to restrict the pointing to a certain set of possible targets. Use of deictic references would probably also require the perspective-taking module, described in Section 2.4.4, because the deictic references are usually given from the speaker’s point of view [29].

Deictic gestures are essentially information about spatial relations and locations. This means that, in addition to the deictic speech utterances, a deictic user interface module capable of using deictic gestures requires only this spatial information as an input in order to define the task unambiguously, as shown in Figure 2.12.

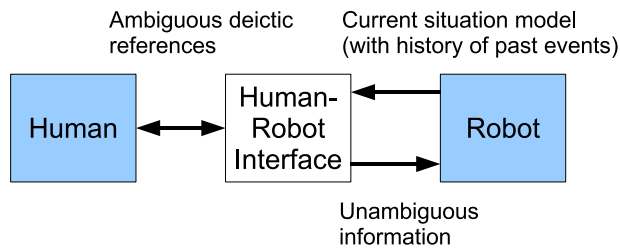


Figure 2.12: Deictic references communication method described as a user interface module.

2.5 Affordances - action possibilities

It is known that humans prefer to communicate with robots on a level where they think the robots will correctly understand them [132]. In most cases, this means short and simple utterances that do not leave room for misinterpretation. This reflects the basic requirement of human-robot task communication, which is that the task request utterances used need to be usable both for the human and the robot [42, 73, 65].

This fundamental requirement is also a starting point for the so-called affordance-based task communication method, which is presented in this thesis. The idea is that only a reference to a task-related target object or action can communicate the whole task for a robot that is capable of associating objects with actions that the robot can perform with those objects.

An example of this kind of indirect task communication is shown in Figure 2.13. Instead of directly communicating all the task parameters, which in this case are the “analyse” action and the “rock” target object, the human can use affordance-based indirect task communication by stating only the task’s action or target object name, i.e. in this case, “analyse” or “rock”. The robot can then complete the task request by using the knowledge of what actions it can perform with the referred object.

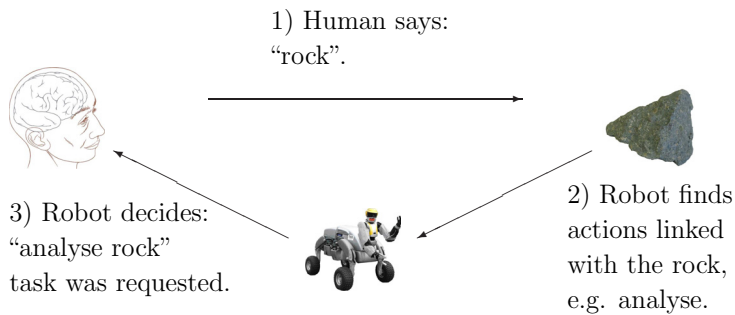


Figure 2.13: Indirect human-robot task communication using only task-related object (or action) reference is possible if the robot can associate objects with actions that it can perform with the objects.

2.5.1 Concept of affordances

The underlying idea researched in this thesis is the usage of known object properties to limit the possible tasks that can be performed with the objects. This approach is based on the so-called “theory of affordances” [45], which postulates that all objects can be considered to have a property called affordance that defines which actions are possible in relation to the actors. Affordances are formally defined in the theory of affordances as “action possibilities in the environment in relation to the action capabilities of an actor”.

The concept of affordances was created and elaborated largely by researchers in the field of psychology. The concept of affordances can be considered to have been derived from the way the perceptual systems of animals evolved from the need to control and guide actions [3]. Thus, the perception of actions is not merely an additional feature of human perception, but is instead the initial reason human perception abilities developed.

Several subsequent studies provided further information on the role of affordances in human cognitive processes. For instance, it has been shown that human perception of objects enables a direct association with the possible actions that can be performed with those objects [47, 135]. This means that, for example, seeing a rock activates in the brain action presentations such as “analyse” and “pick up”.

Furthermore, it has been shown that no other indication of action other than the object itself is required for the object action-association to occur, and that the perception can also be a form other than visual perception for the object-action association to work [47]. In addition, the object does not need to be visible to the human at the time of the action selection [135]. This means that probably almost any type of reference to an object is able to trigger the action presentations in the brain.

Thus, perception of objects alone can convey information about the object-related actions [136]. For example, when a person indicates an exit door to somebody, they also convey implicitly the possible actions, such as “go out” or “open the door”. Object reference alone is thus enough to communicate the whole task consisting of action and target object. Especially when the actor’s action possibilities are quite limited, as is the case with an average service robot, the object-action associations can unambiguously define the desired task.

2.5.2 Affordances in user interface research

The idea that objects and actions are linked has been successfully adopted into use in the field of human action recognition. First of all, the observed human actions have been used to identify objects in the environment [104]. The idea is that certain actions, for instance typing, can be performed with certain objects, such as a keyboard. Thus if we have an *a priori* list of the objects' affordances, we can automatically classify objects just by observing the user actions.

This object-action link has also been used in the other direction, i.e. to recognise actions based on the observed objects [89]. However, the underlying idea is the same: certain actions can be performed only with certain objects. In this case, we also need an *a priori* list of affordances that we can then use to link the observed objects with the possible actions.

The difficulty with this type of object-action association is proportional to the complexity of the environment, i.e. to the number of possible actions and objects in the environment [89]. In particular, actions that can be related to several objects, such as picking up, require additional information, for example from the work context, to make the association unambiguous. This complexity constraint is also applicable when we communicate using affordances. The greater the number of different objects and actions we need to consider, the more likely that communication is ambiguous.

Context menus in graphical user interfaces also utilise a concept similar to affordances. The context menus have long been used to provide context-related menu entries [99]. For example, the context menu entries could be "open" and "delete" actions if a pdf-document object is being selected. This means that a computer is utilising its knowledge of the selected object's affordances, i.e. what actions it can perform related to that object.

2.5.3 Affordances in robotics research

Performing actions on objects is considered to lie at the heart of HRI [73]. The main goal of human-robot communication in particular is to transfer these two linked parameters between the human and robot actors. The concept of affordances has thus long been, in one form or another, at the core of robotics research.

Object-action associations have already been utilised in task communication, although not explicitly based on the concept of affordances, to confirm the validity of a human-to-robot task request [28, 40]. In this case the robot’s language processing also has the ability to complete partial task requests, but this is only done based on previous task utterances and user dialogue, i.e. without utilising the defined object-action associations.

Another similar type of task communication approach instead utilises the associations directly for task communication [66, 77]. The idea is that a human can ask the robot what objects it knows and what actions can be performed with individual objects. This kind of object-action association has also been used to verify the consistency of speech utterances [76]. In both cases, however, the robot does not perform any automatic associations but requires the task parameters to always be communicated explicitly.

One other robotic application, which explicitly utilises the affordances, introduces a way to automatically detect affordances from the robot’s environment [90]. The idea of this robotic subsystem is to scan for spatial relationships from the environment that meet the requirements of certain actions. For example, “chair” is an object that affords the action of sitting. For this functional purpose, a chair has a flat area at a height of a few dozen centimetres from the ground. The chair is thus defined through the actions that it affords. This type of automatic perceiving of affordances could be a valuable counterpart to the communication system described in this thesis.

2.6 Conclusion

Several astronaut assistant robots have been built in order to research astronaut-robot cooperation. The development of these robots started from basic functional requirements such as human-speed mobility, dexterous object manipulation, and environment perception, which are also actively researched for terrestrial robotic applications. However, communication between astronauts and robots became an issue as soon as the robots began to be able to perform tasks autonomously.

The human-robot interfaces relevant to astronaut-robot communication during planetary surface missions are mainly speech and gesture-based, because they can also be used with the bulky space suits used by the astronauts. Speech in particular has been the main interface for communication in all human-inspired astronaut-robot interaction systems that have been developed. Speech is also used with the human-inspired astronaut-robot task communication that is developed in this thesis based on the concept of affordances.

The concept of affordances, i.e. action possibilities, has already been successfully applied, for example, to recognising activities, to validating task requests, and to finding action possibilities from the work environment. However, affordances have not been used to enable indirect task communication using only object or action names, as is done with the proposed affordance-based task communication.

The concept closest to affordance-based task communication is a context-menu that has long been used with GUIs. The idea of a context menu is to generate for the user a list of actions related to a selected object. The user then defines the required task by selecting the correct action from the list. In addition to explicit natural language task requests, this kind of listing of object-related actions, which is not initiated with object or action name utterance alone, is the conventional method against which the newly formulated affordance-based task communication is evaluated in this thesis.

3 Requirements for Robotic Astronaut Assistants

The purpose of this chapter is to learn what kind of tasks astronaut assistant robots might be required to perform with astronauts on planetary surfaces. These tasks define what kind of capabilities will be required from a robotic astronaut assistant, both from the task execution perspective and from the task communication perspective. In particular, the analysis considers how these activities could be performed with a centaur-like outdoor service robot, called the WorkPartner.

The research methodology utilised in this chapter is systematic literature review [98]. The purpose is to identify the most important documents reflecting future ESA and NASA Moon and Mars missions, and to extract from them five common mission scenarios. This is done in Section 3.1. These missions are then further broken down hierarchically into tasks in Section 3.2. In Section 3.3, the missions are further used to define the robotic astronaut assistant capability requirements for performing the required activities. In addition, before concluding the chapter in Section 3.4, the capability requirements are compared with the WorkPartner robot's current capabilities. This makes it possible to point out the concrete technology development efforts required to make the WorkPartner robot into a useful astronaut assistant robot.

3.1 EVA astronaut mission scenarios

The EVA astronaut activity analysis starts by identifying the most common EVA astronaut activities for a surface exploration mission. This identification is done by reviewing the latest, i.e. published between 1998 and 2008, NASA and ESA documents that address manned exploration of the surfaces of the Moon or Mars. The documents were collected from Aalto University and ESA electronic databases.

All the available documents that addressed “manned exploration” of the surfaces of the “Moon” or “Mars” were reviewed. Search criteria limited the selection to only those documents that also examined possible astronaut activities on surface exploration missions. A total of five such documents were identified. There were several other documents that described the objectives of the exploration of the surfaces of the Moon and Mars. Most of them, however, were used as inputs in the five documents and are therefore not described here.

The five surface exploration documents reviewed were

- Lunar Exploration Objectives (LEO), 2006 [93]
- NASA’s Exploration Systems Architecture Study (ESAS), 2005 [122]
- The Mars Surface Reference Mission (MSRM): A Description of Human and Robotic Surface Activities, 2001 [61]
- The Lunar Surface Reference Mission (LSRM): A Description of Human and Robotic Surface Activities, 2003 [30]
- Human Mars Mission Project (HMMP): Human Surface Operations on Mars, 2004 [74]

The first reviewed document tries to list all possible lunar exploration themes (“why we go there”) and objectives (“what we do there”) [93]. The document was published by NASA in December 2006 as a first version of all the objectives that anyone might pursue in lunar exploration. The objectives help to define the required core mission activities, but a set of support activities such as infrastructure setup and facility maintenance must also be included from other documents.

The second reference document presents the results of a 90-day NASA internal study of how to implement NASA’s “Vision for Space Exploration” [97]. It presents

NASA's view of the most likely space exploration architecture and also describes the probable tasks of EVA surface missions.

The third and fourth reference documents are NASA studies that are especially focused on describing the activities that would be performed on the surfaces of the Moon and Mars [61, 30]. Their purpose is to describe "what" activities would be done, rather than "how" they would be done.

The last document reviewed is an ESA technical document that describes astronaut surface operations on Mars [74]. It explains the EVA activities that need to be performed by a Mars surface mission and also describes their time requirements. It provides a non-NASA perspective on the scenario analysis.

Table 3.1 presents all the commonly identified activities in the analysed surface exploration documents. When applicable, the document name, along with a page number or activity code, is mentioned for the activities. These activities can also be seen as mission objectives, defining the purpose of the activities. The five most commonly identified scenarios are: geological exploration; scientific experiment deployment; facility maintenance; communication network setup; and dust removal.

Geological exploration was explicitly defined as one of the mission objectives in all of the reference documents, as shown in Table 3.1. Geological exploration includes exploration and sample measurements. The geological exploration scenario can be divided thus into the following parts: (1) take the required tools for geological field exploration from the storage area, (2) explore a specified area in the environment for interesting samples, (3) collect interesting samples and perform preliminary sample analysis, (4) document all relevant information and store the samples (sample curation), and (5) return the samples and tools to the storage area.

Table 3.1: EVA astronaut planetary surface activities from the ESAS [122], MSRM [61], LSRM [30], LEO [93], and HMMP [74] documents. The page numbers and activity labels can be used to locate the activity descriptions from the documents.

EVA astronaut planetary surface activities	ESAS	MSRM	LSRM	LEO	HMMP
+Geology				mGEO	
Sample collection (surface and subsurface)	p199	p18	p10,p23	mCAS1	p9
Sample storing (curation)	p199	p18	p10	mGEO15	p9
Describe geological relationships	p199	p19	p10	mGEO10	p5
Surface exploration/scouting (kilometers)		p18	p10	mSM1	p9
Emplace geophysical instruments	p199	p19	p10	mGEO3	p9
+Communication				mCOM	
LAN infrastructure	p206	p21	p17	MCOM1.3	
Communication links to Earth	p206		p13	mCOM1.2	p5
+Inspection, maintenance, repair					
Surface facility assembly	p557	p12	p92	mSM3	p9
Surface facility maintenance (check and repair)	p557	p12	p109	mSM3	p9
Logistics (transport supplies for base)	p557	p82	p25	mSM2	p9
Dust mitigation (dust removal)	p557	p32	p19	mEHM2	

The deployment of scientific experiments is also identified as a mission objective in all of the reference documents. The experiments can be geophysical experiments, environment characterisation experiments, or astrophysical experiments, for example. All of these experiments require similar tasks in order to be deployed; only the experiment-specific initialisation procedures differ. The scenario can be thus divided into the following parts: (1) get the experiment package and required tools from storage, (2) explore the environment and identify a suitable location for the experiment, (3) prepare the location for the deployment of the experiment, (4) set up the experiment by following the experiment-specific deployment procedure, (5) document the setup procedure for the experiment, and (6) return the tools and equipment to the storage area.

The Local Area Network (LAN) setup activity was mentioned in all the other documents but not in HMMP [74]. The LAN provides a means of communication on the planetary surface between habitats, astronauts, robots, and rovers. The LAN infrastructure is primarily set up in the areas where the mission activity is located. Modifications to the LAN infrastructure might also be required if activity in a certain area is finished and activity has started in a new area. The LAN setup scenario can be broken thus down into the following tasks: (1) get the LAN base stations and tools from storage, (2) find the exact installation locations in the selected deployment areas, (3) install the base stations in the selected locations, and (4) return the tools to the storage area.

The need for facility maintenance on planetary surface exploration missions was mentioned in all of the analysed documents. Maintenance includes both periodic checks on the facilities and repairs to the facilities. Facility maintenance is crucial for all types of missions in order to guarantee crew safety in hazardous planetary surface environments. The facility maintenance scenario can be broken down into the following tasks: (1) check the facility to identify the repair needs, (2) get the required tools from storage, (3) carry out the repair procedures, and (4) return the tools to storage.

The dust removal activity was mentioned in all the other documents examined but not in HMMP [74]. The dust removal activity includes removing dust from equipment, facilities and from EVA astronaut space suits. Dust can cause health risks for astronauts, and reduce device performance. The dust removal scenario can be thus broken down into the following tasks: (1) get the required tools from the storage area, (2) identify the areas that need to be cleaned, (3) use the tools to clean the area, (4) document the cleaning activity performed and its results, and (5) return the used tools to the storage area.

3.2 Mission scenario breakdown

The second step in the activity analysis is to break down the defined missions into tasks. The idea is to find the minimal set of tasks that are required to build the five most typical missions. The mission scenario breakdown and analysis is performed using the ESA Control Development Methodology (CDM) [107, 33]. Another similar alternative could have been the Work Domain Analysis (WDA) [95].

The idea of CDM is to provide traceability between requirements and final realisation by clearly indicating when constraints are laid down and engineering design decisions are made. CDM principles can be seen as principles for writing good requirements. Only the first phase of CDM, i.e. activity script definition, is utilised here. The activity script analyses in detail missions, tasks, and subtasks, i.e. activities, and it can be further used to conceive a system architecture for performing these activities.

The CDM tasks are extracted based on the overall mission descriptions given in Section 3.1. An example of this kind of task extraction is shown in Table A.1 for the LAN setup mission. All the CDM tasks extracted from the five identified mission scenarios are shown in Table 3.2. The number under the mission heading indicates how many times the tasks were needed in each of the missions. Additionally, the tasks that can be run at any point during the mission, or that can be run parallel to the main mission, are listed in the last column of the Table 3.2. This kind of parallel task is for example mission progress monitoring.

The most commonly used tasks are moving to a new location (TRANSPORT), relocating objects (RELOCATE), and providing information on the environment (INSPECT). The rest of the tasks, i.e. the loading and unloading of tools (LOAD/UNLOAD), performing complex automated processes (PROCESS), and defining mission parameters (DEFINE), are all also required in at least three different missions. All the tasks except DEFINE are mentioned in the CDM document [107, 33]. The

DEFINE task was not required in the CDM document because the missions, situated in the relatively static orbital space environment, were assumed to be initially properly defined and not require any online modifications.

Table 3.2: List of CDM tasks used in the five mission scenarios and in the tasks available in parallel during the missions.

CDM task	Task description	Geological exploration	Experiment deployment	Dust removal	Facility maintenance	LAN setup	Parallel tasks
TRANSPORT	Move to a new destination	3	7	4	8	4	1
RELOCATE	Transfer object to new location	4	7	3	9	4	0
INSPECT	Provide surveillance of a scene	4	12	4	15	5	4
LOAD	Prepare a tool for operation	2	0	1	0	1	0
UNLOAD	Undo the effect of LOAD	1	0	1	0	1	0
PROCESS	Invoke a complex automated process	1	1	1	1	1	0
DEFINE	Determine parameters for a mission	1	1	2	1	2	6

The seven different tasks shown in Table 3.2 are further divided into 17 subtasks. These 17 subtasks are shown in Table 3.3. An example of how the tasks are divided into subtasks is shown in Table A.2. The numbers in Table 3.3 indicate how many times the tasks are used in each of the subtasks. The most commonly used subtasks are the calculation of new state values (EVALUATE), sending information to other systems (SEND), and measuring process values (MEASURE). They can be seen as the most important building blocks of a mission, without which none of the tasks could be performed. The second-most commonly used subtasks are the manipulation-related APPROACH, EXTRACT, and INSERT. There are also five subtasks that are required only for one task each.

Table 3.3: List of all subtasks used in the seven CDM tasks. The numbers indicate how many times the task uses the subtask.

Subtask	Subtask description	TRANSPORT	RELOCATE	INSPECT	LOAD	UNLOAD	PROCESS	DEFINE
ACQUIRE	Acquire system internal state information	2						1
ACTIVATE	Activate a device						1	
ADJUST	Set a device state to a value				1			
APPROACH	Position subject with target without contact		1	1	1	1	1	
ATTACH	Establish rigid connection		1		1			
DEACTIVATE	Undo the effect of ACTIVATE						1	
DETACH	Undo the effect of ATTACH		1			1		
DISPLACE	Move to a goal pose along any path				1			
EVALUATE	Compute a state information	4	7	6	3	3	2	1
EXTRACT	Undo the effect of INSERT		2		1	1	1	
FOLLOW	Move subject, e.g. tool, along a path		1					
INSERT	Place subject within confinement of a target		1			1	1	
MOVE	Position subject to a goal pose along a path	1		1				
MEASURE	Acquire state information	1	2	2	2	2		2
RETRACT	Undo the effect of APPROACH					1	1	
SEND	Send a message to another actor, e.g. robot	4	4	5	2	2	2	1

3.3 Robotic astronaut assistant requirements

Next, the defined activity script is used to define the requirements for the robotic astronaut assistant capabilities needed to perform the defined activities. The required capabilities for all of the examined mission scenarios are very similar, as can be seen from Table 3.2. There is, for example, a common need to move autonomously, recognise objects, and monitor the progress of the mission scenario.

Each one of these tasks poses requirements that the astronaut assistant robot has to meet in order to be able to perform the tasks, as shown in Table A.1. The tasks can thus be mapped to corresponding technology development requirements for the astronaut assistant robot.

The identified technology development requirements are shown in Figure 3.1. They can be grouped into three different research areas: shared situation awareness, task coordination, and robot action control architecture. The goal of shared situation awareness is to provide a shared understanding of the information relevant to the situation. Task coordination, on the other hand, aims to define performable missions and provide a means to solve unexpected events during nominal mission performance. Finally, robot action control enables the robot to move and manipulate its environment.

Some of the defined capability requirements fall into more than one of these groups. For example, the semantic information dialogue can be used both for providing situation awareness and for solving unexpected events. The main purpose of the categorisation is to provide an understanding of the high-level goals towards which the individual requirements contribute.

3.3.1 WorkPartner robot

Aalto University's WorkPartner robot, shown in Figure 1.1, has been in the process of development for a decade now; the robot is intended to facilitate cooperative task performance with humans. Its initial designated work domain was light outdoor tasks such as garden work (picking up and moving objects, blowing snow) and light forestry tasks (cutting trees, piling up objects). It is designed to work as an interactive partner by using interfaces that would enable natural and seamless cooperation in task performance. Next, the WorkPartner robot will be utilised and

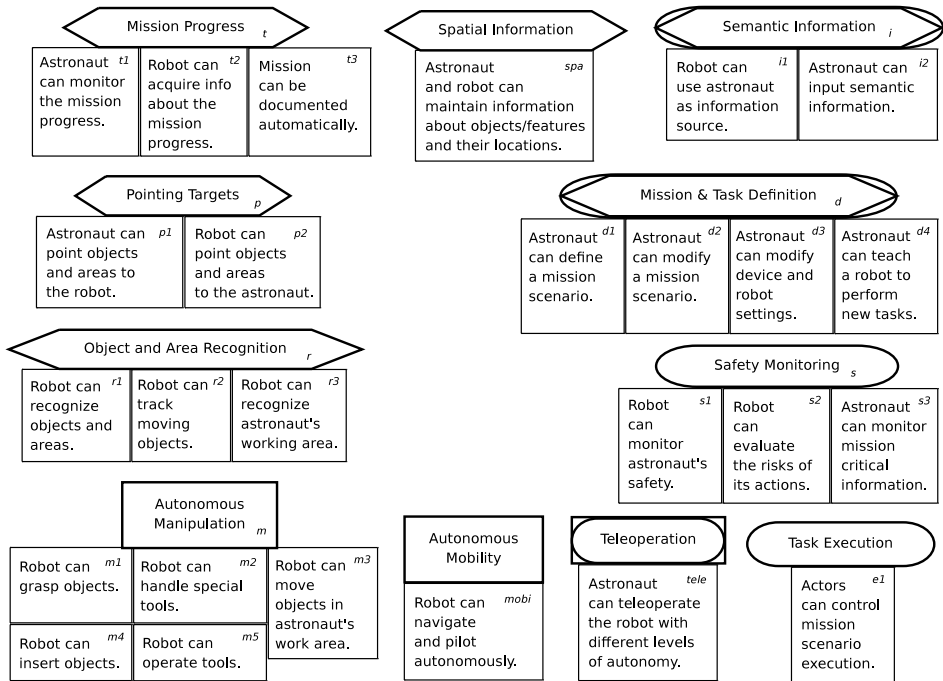


Figure 3.1: Requirements for robotic astronaut assistant capabilities on surface exploration missions: shared situation awareness (arrow box), task coordination (circular box) and robot action control (rectangular box).

further developed in order to make it capable of performing as a robotic astronaut assistant.

Currently, the WorkPartner's most important technological capabilities as an astronaut assistant are its four-legged wheel-walking-based mobility, two-arm gripper-armed manipulation, multimodal human-robot interfaces, modular task definition architecture, autonomous navigation, and object recognition and tracking [70, 129]. Thanks to these capabilities the WorkPartner robot can already perform several tasks that might be required on space exploration missions. The WorkPartner can, for example, follow an astronaut, pick up items that are pointed out to it, and navigate autonomously across various known and unknown terrains.

3.3.2 WorkPartner readiness

Last, the readiness of the Aalto University WorkPartner robot to meet the capability requirements can be evaluated. The readiness of the WorkPartner in year 2008 to meet the requirements described in Figure 3.1 is shown in Table 3.4. The table shows that the WorkPartner has some readiness to meet all of the identified capability requirements. The WorkPartner’s strengths currently lie in tele-operation and autonomous mobility. The technological capabilities for modifying the defined missions and sharing semantic information between robotic and human actors, on the other hand, require more development in order to be useful.

Table 3.4: WorkPartner readiness to meet capability requirements of an astronaut assistant robot is rated from one (bad) to five (excellent). The “Id” column refers to Figure 3.1.

Id	WorkPartner readiness	Id	WorkPartner readiness
r1	2, only specific objects are recognised	d3	2, only robot settings can be modified
r2	3, only specific objects are tracked	d4	2, new tasks are programmed manually
r3	3, human located by laser scanner	s1	1, only some human action recognition
p1	3, using pointing stick and laser pointer	s2	2, positioning relative to the robot
p2	4, using arms and laser pointer	s3	3, robot status displayed to human
t1	2, mission progress of robots available	e1	2, execution start, pause, and stop
t2	3, robotic actor’s task progress available	tele	3, tele-operation interface exists
t3	3, robot current task progress displayed	mobi	4, using laser scanner-based navigation
spa	2, only specific objects are understood	m1	3, only specific objects can be grasped
i1	4, bi-directional queries supported	m2	2, end effectors changed manually
i2	1, only raw audio recording available	m3	2, using low-speed actions with human
d1	4, mission scenario builder exists	m4	2, only specific objects can be inserted
d2	2, runtime modifications very limited	m5	3, tool operation definition possible

3.4 Conclusion

This chapter presented an assessment of the technological requirements for a robotic astronaut assistant such as the centaur-like outdoor service robot of Aalto University, named the WorkPartner. The chapter analysed five documents dealing with missions to the surfaces of the Moon or Mars and extracted from them the most probable surface activities that would involve an EVA astronaut. Five mission scenarios were constructed from these surface activities: geological exploration; scientific experiment deployment; dust removal; facility maintenance; and local area network setup mission scenarios.

Five EVA astronaut mission scenarios were analysed by breaking down the missions into tasks and the tasks into subtasks. The tasks and subtasks were then used to identify 26 technological capabilities required by robotic astronaut assistants. These technological capabilities were broadly divided into three technology frameworks: shared situation awareness, task coordination, and robot action control.

The shared situation awareness framework provides an understanding of the environment, tasks, and actors. The task coordination framework utilises this information to decide if missions can be performed, and it also provides a means of solving unexpected events during the nominal performance of the mission. Finally, robot action control enables the robot to move and manipulate its environment.

The WorkPartner robot has some readiness concerning all the technologies identified above, but some of the technologies still require further development if the robot is to be truly useful to astronauts. WorkPartner robot technologies are most mature in the areas of tele-operation and autonomous mobility. The sharing of human and robot information and on-the-spot modifications to mission scenarios require more development to be truly usable.

In conclusion, the WorkPartner robot in its current version could already perform planetary exploration missions, such as geological exploration or facility maintenance. The next challenge is to further develop the WorkPartner's capabilities. The focus of this thesis is human-robot task communication, which was part of the capabilities posited to require special attention. This selected focus is examined in the next chapters with user experiments constructed using the tasks identified here as typical of future astronaut-robot missions.

4 Unambiguous Task Communication Using Affordances

The purpose of this chapter is to examine how efficiently the affordance-based task communication method can be used to communicate tasks to the robot in the field. The aim is to get a preliminary idea of the potential of the approach by restricting the environment to a case where each object is unambiguously associated with one action, and vice versa.

Section 4.1 first presents a geological exploration mission where participants request the robot analyse rocks and set up measurement units. The experiment is performed with a fully autonomous mobile WorkPartner robot. Section 4.2 continues the first experiment with a mission scenario where the astronaut is required to solve emerging problems. This second experiment is likewise performed with a fully autonomous WorkPartner robot, but this time the robot platform is static and only the robot manipulators move during the experiment.

4.1 Geological exploration experiment

The question examined in this section is whether affordance-based indirect task communication can be used to improve human-robot task communication in an operating environment where only unambiguous object-action associations exist. This unambiguous situation is examined with a user experiment performed with a fully autonomous centauroid robot in a geological exploration mission context.

4.1.1 Method

Participants

A total of 12 participants were selected for the experiment. Nine of the participants were male and three were female. The average age of the participants was 28.0 ± 3.3 years. All participants, except for one law student, were Aalto University students. None of the participants spoke English as their native language. All of the participants can be considered novices, since they did not have any previous experience with the examined system. Participants were compensated for their participation with a movie ticket.

Equipment and software

The experiment was performed in a large indoor lobby area, shown in Figure 4.1. The usable test area was approximately 8 metres wide and 30 metres long.

Aalto University's WorkPartner robot [49], shown in Figure 4.2, was used as a fully autonomous astronaut assistant robot in this experiment. The robot moved by using its four wheels and by utilising a middle platform joint for turning. The SICK LMS291 laser rangefinder attached at the WorkPartner's waist was used to track the participants' locations, and to enhance wheel odometry-based localisation by matching consecutive scans to calculate the robot's movement relative to the environment. The robot torso has two DOF at the waist, enabling the torso to tilt and rotate, and five DOF in both the right and left arms. The WorkPartner's head is mounted on top of a two DOF pan-tilt unit, and has an LED array used to animate the mouth movements when the robot speaks.



Figure 4.1: The experiment area. The sheets of paper on the ground were used to cover the rocks in order to make it more difficult for the participant to locate interesting rocks.



Figure 4.2: The WorkPartner robot worked as an astronaut assistant robot in the experiment.

The constraints created by an astronaut's space suit were taken into consideration by using restrictive clothing, shown in Figure 4.3. This clothing consisted of a heavy backpack, wooden platform sandals, and a helmet. The purpose of the clothing was

to restrict the participants' movements so that, for example, picking up items from the ground was a difficult task. The participants had also a Shure PG1 wireless microphone attached to their chest for speech communication.



Figure 4.3: The restrictive outfit used to mimic astronaut space suit constraints consisted of wooden platform sandals, a backpack and a helmet.

The participants and the WorkPartner robot operated with three different types of objects in the experiment, as shown in Figure 4.4. Only the red rocks were considered interesting to the astronaut. The covered rocks were laid out in the environment in random pairs, with no particular attention being paid to the composition of the pair, i.e. red rocks and normal rocks were randomly chosen for the pairs. The idea behind this arrangement of pairs was to require the participants to take interest in observing the rocks, rather than automatically assuming that the rocks were interesting. Initially, the sheets of paper were also designed to detect the exact moment when the participant found an interesting rock, but in the end, this information was not utilised due to its inaccuracy.

The measurement unit mock-up was an empty cardboard box covered with metal-grey tape. The measurement unit was carried on the WorkPartner robot's back from where it was picked up by the robot and placed at the ground location requested by the participant.



Figure 4.4: Measurement unit mock-up (above) and a few interesting (lower left) and uninteresting (lower right) rocks used in the experiment.

Two dual-core laptop computers were used in the experiment. The first one was on the top of the WorkPartner robot, and it took care of performing the requested tasks, such as following the participant. The second laptop was on a table next to the experiment area and it was used to receive sound signals from the wireless microphone, and to run the software for speech recognition and for the dialogue manager. These two laptops communicated with each other through wireless LAN.

The speech recognition software used in the experiment was CMU Sphinx II [64]. This software output recognised words, which were then sent to the dialogue manager, as shown in Figure 4.5, to be interpreted using a frame-based approach [84], i.e. a task request is accepted when all the task-related parameter slots have been filled. The software architecture used in the experiment is described in Section B.2.

The WorkPartner robot was able to perform five different tasks in this experiment, corresponding to five task requests. The first one was the “stop” task, which stopped all robot movement (both the wheel and manipulator). Another similar speech request was “wopa”, the nickname of the robot. It stopped the wheel movements and drove the manipulators to a zero position, which is the manipulator configuration shown in Figure 4.2. The robot head, however, faced the participant in all situations

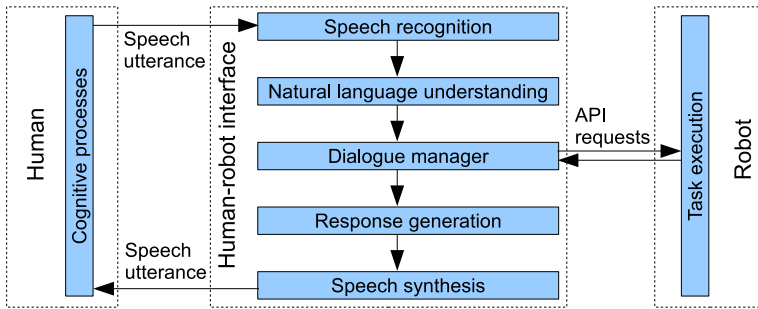


Figure 4.5: High-level architecture diagram of the astronaut-robot cooperation system.

except when “stop” was requested. The robot located the participant by tracking a group of 2D points from the WorkPartner’s rangefinder data. The participant was able to recover lost tracking by going to an area between the WorkPartner’s manipulators.

The third task was “follow”, which enabled the WorkPartner to maintain a safe distance of two metres from the participant. This was done by calculating a new target navigation point for the robot that was two metres away from the participant in the direction of the robot. The robot was able to drive to any given coordinate that was directly reachable with the WorkPartner’s four-metre turning radius.

The two remaining tasks were the “analyse rock” and “set up unit” tasks. When the “analyse rock” task was requested, the WorkPartner robot drove to analyse the rock in the location where the participant was standing at the moment the request was made. In this way, the participant’s 2D position worked as a location-pointing interface. The rock analysis consisted of having the robot bend over the rock and move its left hand back and forth a few centimetres above the rock. Since the hand had no actual sensors to be utilised for analysis, the rock was then randomly designated as interesting or uninteresting, with a 90% chance of being interesting.

In the “set up unit” task, the WorkPartner grasped the on-board measurement unit with the two-finger gripper attached to its right arm and placed the unit on the ground at its own location.

Speech utterances and the participant’s physical location were the only two interfaces through which the participant could communicate with the robot. The speech interface was always used to give the overall task parameters, which were supplemented with the participant’s location information in the “follow”, “analyse rock” and “set up unit” tasks.

The robot did not have any a priori knowledge of objects in the environment, so utterances referring to objects had to be accompanied by position information. This position information was the physical location of the participant at the moment of the task request. The robot was unable to recognise any objects in the environment except for the participant.

Robot-to-human communication was done through Festival² speech synthesis software, aided by the robot’s head orientation and mouth expressions, along with the robot’s location and orientation. The main communication method was speech, through which the robot acknowledged all the participant’s task requests by describing what the robot planned to start doing next.

Three different speech-based task communication methods were defined for requesting tasks from the robot. With the first task communication method, called the *action with object* or *direct* task communication method, all of the task parameters, i.e. action and the target object, had to always be communicated explicitly in the request. The second and third communication methods, namely *affordance-based* or *indirect* task communication methods, were based on the concept of affordances presented in Section 2.5. With this approach the object-related action possibilities

²<http://www.cstr.ed.ac.uk/projects/festival/>

were utilised by the robot to complete the requested task. For example, based on a rock-analyse object-action association, the robot can derive the task to be “analyse rock” when the human communicates only the object name “rock” or the action name “analyse”. Direct association of action or object names to tasks is possible in unambiguous cases because each object is associated with only one action, and vice versa.

All the possible tasks and the corresponding task request utterances are listed in Table 4.1. The dialogue structures of the task communication methods are shown in Figure 4.6. The implementation differences between the direct action with object and the indirect affordance-based methods are shown in Figure 4.7. The main difference in implementation is that, in the affordance-based method, the robot’s database of known object-action associations is also used to interpret the action or object names as task requests. This object-action database can be learned automatically based on the direct task requests or, as in this case, be defined a priori.

Table 4.1: All the possible tasks and task requests in the unambiguous user experiment. The affordance-based task requests are marked*.

Task description	Action with object	Action*	Object*
Analyse a rock on the ground	Analyse (the) rock	Analyse	Rock
Setup a measurement unit on the ground	Set up (the) unit	Set up	Unit
Stop all robot movement	Stop	Stop	Stop
Request robot’s attention	Wopa	Wopa	Wopa
Request the WorkPartner to follow	Follow	Follow	Follow

Experimental design

The experiment used a repeated measures, i.e. within subjects, experimental design with one independent variable and two dependent variables. The independent vari-

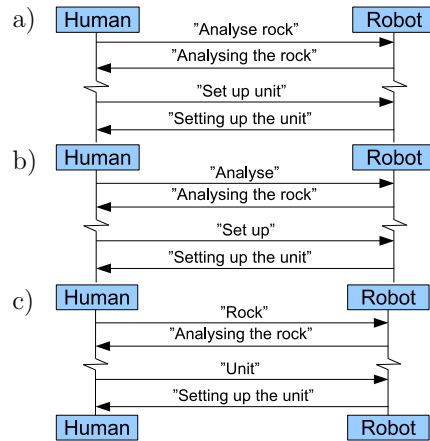


Figure 4.6: The dialogue structures of the three compared task communication methods. The *direct* task communication method (a) requires all task parameters to be communicated explicitly, whereas the *indirect* affordance-based task communication methods require only the action name (b) or object name (c) to be communicated.

able was the communication method with three different levels: direct action with object, action name, and object name. These three task communication methods were presented in the previous section. The two dependent variables were the participants' task communication workload and their task communication preferences.

The experiment also included a qualitative assessment part. The goal of qualitative assessment was to observe how participants work with the examined task communication methods, and what the participants consider to be the strengths and weaknesses of the examined system.

The experiment was counterbalanced in order to eliminate the effect of the order in which the task communication methods were used. There were six possible test round combinations, since there were three levels of the independent variable, i.e. three task communication methods to be tested for each participant. This means that only every sixth participant performed the experiment in the same order.

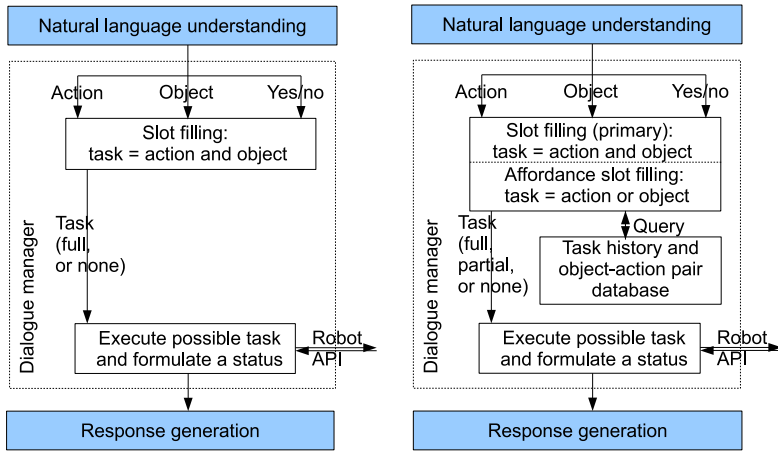


Figure 4.7: Dialogue manager subsystems for the compared direct action with object (left) and the indirect affordance-based (right) task communication methods.

Procedure

The overall mission scenario in the experiment was astronaut-robot geological exploration done on the surface of Mars. The participant was an astronaut working with the WorkPartner robot. The robot followed the participant and performed different tasks based on the astronaut's requests.

The experiment consisted of four different test rounds, which were each performed once by each of the participants. The first three test rounds were identical except that the independent variable, i.e. the task communication method, was changed for each test round. In the fourth test round, all three task communication methods were available for use at the same time.

The order of the first three test rounds was varied so that the same sequence of test rounds was available only for each sixth participant. None of the test rounds was repeated by any of the participants. It was not deemed necessary to repeat

the test rounds, since the participants repeatedly used the task communication methods during the test rounds and they were also able to rehearse using the task communications methods in advance for as long as they wanted.

It can be argued that the relatively short training received by the participants before the experiment is also relevant with regard to real astronaut missions. Although astronauts are well trained, their task performance still relies heavily on the detailed procedures they follow on space missions [9, 83]. On future Moon and Mars missions, the amount of training and experience is expected to decrease because the duration, distance, and complexity of the missions will increase [7]. Other factors such as high workload and microgravity have likewise been hypothesised to impair the performance of astronauts in space [67]. This means that the user interfaces provided to the astronauts must also be usable with minimal experience and training.

The goal for each of these four test rounds was the same: to set up two measurement units next to two different interesting rocks. This means that the dialogue shown in Table 4.2 had to be repeated twice for each of the four test rounds. Each of the participants thus installed a total of eight measurement units next to the eight interesting rocks found in the experiment. Figure 4.8 shows the complete progress of the experiment for one of the participants. The progress of the experiment is drawn on a map generated by the robot during the experiment according to its rangefinder measurements.

The progress of the experiment was as follows for each of the participants: After hearing a description of the experiment's overall mission scenario, the participant was taught to communicate with the robot using speech. The CMU Sphinx 2 speech recognition software was trained separately for each participant by having the participant repeat the words used in the experiment. The participants were told to first practise speech communication in front of a laptop until they were confident that all their utterances were correctly recognised. Then the participants requested all of

Table 4.2: The task communication dialogues used in the experiment. H→R refers to human-to-robot communication and R→H to robot-to-human communication.

Description of event	Direct (action with object)	Indirect (action)	Indirect (object)
1) Get the robot's attention	H→R: Wopa R→H: Yes	H→R: Wopa R→H: Yes	H→R: Wopa R→H: Yes
2) Request the robot to follow	H→R: Follow R→H: Following the astronaut	H→R: Follow R→H: Following the astronaut	H→R: Follow R→H: Following the astronaut
3) Request the robot to analyse a rock, which turns out to be uninteresting	H→R: Analyse rock R→H: Analysing the rock R→H: The sample is not interesting	H→R: Analyse R→H: Analysing the rock R→H: The sample is not interesting	H→R: Rock R→H: Analysing the rock R→H: The sample is not interesting
4) As the rock is uninteresting, request the robot to continue following	H→R: Follow R→H: Following the astronaut	H→R: Follow R→H: Following the astronaut	H→R: Follow R→H: Following the astronaut
5) Request the robot to analyse a rock, which turns out to be interesting	H→R: Analyse the rock R→H: Analysing the rock R→H: The sample is interesting	H→R: Analyse R→H: Analysing the rock R→H: The sample is interesting	H→R: Rock R→H: Analysing the rock R→H: The sample is interesting
6) As the rock is interesting, request the robot to place a measurement unit	H→R: Setup unit R→H: Setting up the unit	H→R: Setup R→H: Setting up the unit	H→R: Unit R→H: Setting up the unit

the five possible tasks from the WorkPartner robot at least once. The participants were free to choose which task requests to use with the robot in this rehearsal phase.

The actual experiment phase started when the participant moved to the previously unexplored experiment area. In each of the four test rounds the participant had to

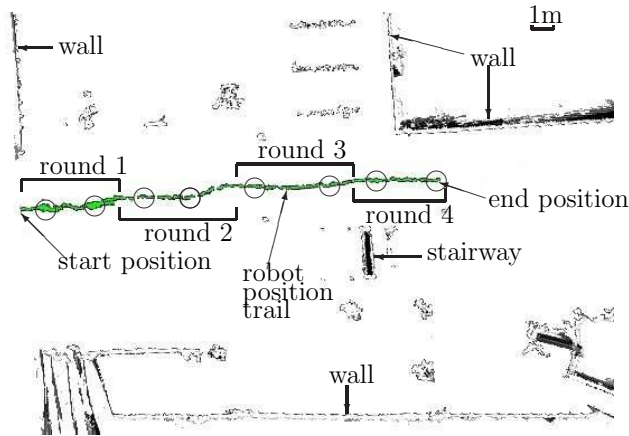


Figure 4.8: Complete progress of the experiment for one of the participants. The progress of the experiment is drawn on a map generated from rangefinder measurements. The circles show the locations where interesting rocks were found, and consequently, measurement units were installed.

perform the following two tasks twice: In the first task, the human had to locate an interesting red rock under the sheets of paper and request the robot analyse the rock. The robot’s analysis was required in order to learn whether the rock actually had scientifically interesting properties. The first task was not completed until the robot reported that an interesting rock sample had been found. The second task was to set up the measurement unit on the ground next to the rock that was confirmed as “interesting”.

All four test rounds were carried out in succession with no pauses. After each test round, the participants were told which communication method was to be used next. In the fourth round, the participants were instructed to choose the communication method that they preferred to use. The introduction and rehearsal phase took approximately 45 minutes, and the four test rounds and the final questionnaire took another 45 minutes.

Data processing and statistical analysis

The participants' subjective assessments of the workload were measured using the NASA TLX [52]. The NASA TLX rating presents a subjective workload score ranging from 0 to 100, from no workload to full workload, respectively. This score is calculated as a weighted average of six workload components: performance, effort, frustration, and mental, physical, and temporal demands.

Three NASA TLX rating sheets were collected from each of the participants, one for each of the three examined communication methods, after they had completed all of the four test rounds. The participants were asked to evaluate only the workload induced by the task communication from the point when they noticed that a certain task had to be done to the point when they started their speech utterance. This means that, for instance, speech recognition performance was excluded from the evaluation.

The number of times that the participants chose to utilise each of the compared communication methods was also counted in the fourth test round. This counting was done manually during the experiment and confirmed later from video recordings and from the robot's log files.

At the end of the experiment, participants answered a questionnaire containing free form and multiple choice questions. This questionnaire was conducted as a contextual inquiry [62] interview, the purpose of which is to treat the participants as experts from whom the interviewer is learning about use of the system directly in the work context. This interview was not done during the test rounds, but after them, in order to not affect the other quantitative performance measurements.

The statistical significance of the results obtained was calculated with R software [108] using the one-way within-subjects ANOVA test. The ANOVA input data

sphericity assumption was checked with Mauchly’s sphericity test. Finally, three post hoc pairwise comparisons were calculated with Bonferroni adjusted paired t-tests in order to identify which specific results differed. A p value of less than 0.05 was the standard for significance.

4.1.2 Results

All of the participants eventually managed to correctly request all of the tasks that needed to be accomplished in each of the test rounds. However, one participant once said “analyse rock” when he was supposed to say “set up unit”. However, the participant noticed the mistake immediately, stopped the robot and communicated the correct task request.

The collected NASA TLX workload measurements are shown in Figure 4.9. The one-way within-subjects ANOVA showed that there was a significant difference between the NASA TLX results of the compared three communication methods, i.e. $F(2,22)=8.01$, $p=0.002$. Mauchly’s test indicated that the assumption of sphericity was not violated (chi-square = 4.70, $p=0.095$). The Bonferroni adjusted pairwise t-test comparison showed that the difference between the direct *action with object*-based task communication and indirect *object name*-based communication was significant ($p=0.030$). Similarly, the difference was significant between the direct *action with object*-based communication and indirect *action name*-based communication methods ($p=0.041$). There was, however, no significant difference between the indirect *action name*-based and *object name*-based communication methods.

The participants’ choice of communication method in test round four did not, however, reveal any significant differences. The direct *action with object*-based task communication was used a total of 11 times, the indirect *action name*-based communication 21 times, and the indirect *object name*-based communication 16 times.

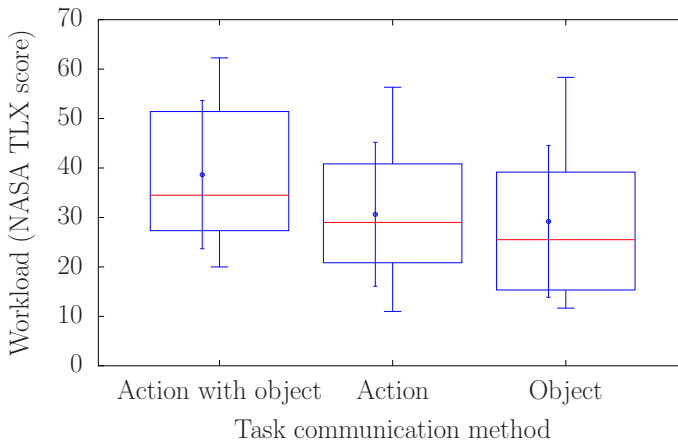


Figure 4.9: Boxplot of the NASA-TLX workload for the three compared task communication methods. Workload values range from 0 to 100, i.e. from no workload to full workload, respectively. Means and standard deviations are shown on the left sides of the boxplots.

A post hoc comparison between direct and indirect task communication methods was also performed in order to see if indirect affordance-based task communication was used more than direct task communication. This comparison can be seen in Figure 4.10. The one-way within-subjects ANOVA showed that the difference between the averages is significant $F(1,11)=10.39, p=0.008$.

Answers to the final user questionnaire provided insights about the potential advantages and disadvantages of affordance-based task communication. The participants commented that indirect *object name*-based task communication was an obvious way to restrict the task to only one place and to make it easy to remember because the required speech utterance was the name of an already visible object. In comparison, indirect *action name*-based task communication was deemed to be a natural way to request something because the utterance is a verb.

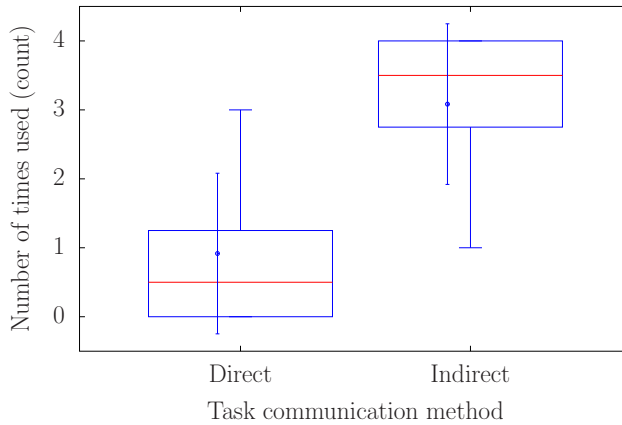


Figure 4.10: Boxplot showing how many times the participants used the direct *action with object*-based task communication and the indirect *affordance*-based task communication methods. The maximum number of task communications is four because there were four tasks that had to be communicated to the robot in the fourth test round.

The advantage of the direct *action with object*-based communication method, compared with the *affordance*-based task communication methods, was deemed to be that it always defines the task request unambiguously. Without full task communication, the robot could accidentally initiate dangerous tasks. This comment was made even though there were no ambiguous object-action associations in this experiment.

4.1.3 Discussion

The NASA TLX results in Figure 4.9 showed that the participants perceived less workload when communicating with *affordance*-based methods, and that they also preferred these methods over direct task communication. Answers to the end questionnaire offered certain possible logical explanations for these findings. The experiment did not, however, directly reveal which, if any, of the proposed explanations could be responsible for the decreased workload.

We know that with the affordance-based task communication method, it is enough to remember only the action or object name related to the task in order to communicate the task request. However, it was not known if communicating with these names would decrease or increase the workload, because people might, for instance, find it more demanding to consider all the possible object-related actions, or conversely, the action-related objects. The workload was nevertheless lower with the affordance-based task communication methods, which indicates that the participants were able to request the tasks without needing to focus intensively on the other possible task-related objects or actions.

The participants were thus able to use the affordance-based task communication methods to provide only the object or action name related to the task without constructing the complete explicit task request. In this way, the affordance-based task communication method probably transferred the cognitive object-action association and task request formulation processes from the human to the robot, which was enough to decrease the workload even in this restricted operating domain.

This type of object-action association process is most likely easier for the robot than for the human. In particular, applying affordance-based task communication to the other direction, i.e. from the robot to human, would probably not help decrease the human workload, but quite the opposite. The affordance-based task communication method is probably essentially easier only for the person making the request and not for the person receiving it. For instance, a surgeon focused on operating on a patient might use affordance-based task communication, for example, by stating the word “scissors” or “adrenaline”, to ask the assistant to perform a certain task, because it is probably faster and induces less workload for the surgeon than the explicit natural language that is normally used. It is, however, very probable that the assistant’s workload is increased, because the assistant has to decide what is the most likely task being requested.

The main contribution of this first preliminary affordance-based task communication experiment was the indication that people can give logical reasons for communicating tasks with action and object names. The two formulated affordance-based task communication methods were not only viable alternative methods, but actually induced less workload. This provides a motivation to further explore the use of affordances in different types of work environments.

4.2 Object manipulation experiment

The first experiment done with the affordance-based task communication method indicated that the method is able to decrease subjective participant workload. The presented experiment extends these results by increasing the number of participants and by measuring objective task communication performance using task communication times.

4.2.1 Method

Participants

A total of 16 participants took part in the experiment. Three of the participants were female and 13 were male. The average age of the participants was 29.2 ± 5.8 years. All the participants, except for one high school trainee, were Aalto University staff or students. None of the participants spoke English as their native language. However, all of them were unfamiliar with the system tested and can therefore be considered novice users. Participants were compensated for their participation with a movie ticket.

Equipment and software

The physical configuration of the experiment is shown in Figure 4.11. The participant and the WorkPartner robot were situated next to a lander mock-up, which had a radio transmitter, solar panel, and measurement unit on top of it, out of easy manipulation range of the participant.

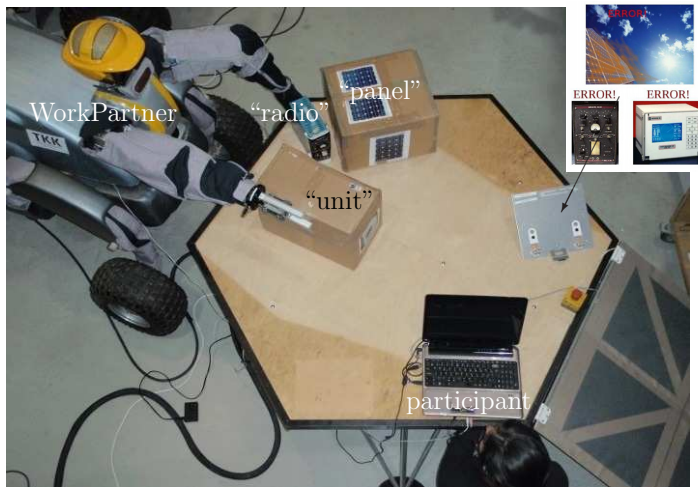


Figure 4.11: The physical setup of the second user experiment. The WorkPartner robot and the participant are located next to a lander mock-up having three items on top of it: a solar panel, a radio, and a measurement unit. Emerging problems were displayed on sheets of paper displayed on the stand in the top right corner.

The WorkPartner robot did not move its platform during the experiment, but performed all the tasks from one location. The participant sat in front of the lander during the whole experiment. A headset was the only additional equipment the participant wore during the experiment.

The measurement unit and solar panel objects were represented in the experiment by cardboard boxes covered with colour figures representing the objects. A voltage meter was used to represent the radio transmitter.

A stand, shown in the top right corner in Figure 4.11, was used to display the emerging problems that needed to be solved by the participants. The problems were shown on the stand on three sheets of paper, each sheet corresponding to one of three possible problems. Each of the sheets of paper had a picture of the problem-related object and a text “error” on top of the object, as shown in the top right corner in Figure 4.11.

Two dual-core laptop computers were used in the experiment. The first one was on the top of the WorkPartner robot and it took care of performing the requested tasks. The second laptop was on top of the lander and it was used to display multiplications to be completed by the participant and to run the software for speech recognition and for the dialogue manager. These two laptops communicated with each other through wireless LAN. The multiplication task displayed on the laptop is a commonly used secondary task in user studies [44].

The speech recognition software used in the experiment was CMU Sphinx II, as in the first experiment. The recognised words were also processed into task requests with the same frame-based dialogue manager shown in Figure 4.5. The software architecture was also the same, i.e. the one described in Section B.2.

The WorkPartner robot was able to perform three different tasks in this experiment, corresponding to three task requests, namely “reset radio”, “clean panel”, and “take unit” tasks. These tasks are listed in Table 4.3.

When the “reset radio” task was requested, the WorkPartner robot moved its left arm behind the radio mock-up and pushed it gently. The idea of this task was to imitate pressing a reset button on radio equipment. In the “clean panel” task, the WorkPartner utilised a brush, attached to the back of its left hand, to sweep clean the top of the solar panel. In the “take unit” task, the robot grasped the measurement unit with both manipulators and lifted it up in the air. The purpose

Table 4.3: List of all possible tasks, i.e. actions with different objects, in the experiment.

Task	Object	Action	Task description
1	Radio	Reset	Robot pushes the reset button of the radio
2	Panel	Clean	Robot uses a brush to clean the top of the solar panel
3	Unit	Take	Robot lifts up the measurement unit in order to remove a sample that is stuck at the bottom of the unit

of this lifting was to remove a sample that had gotten stuck in the bottom of the measurement unit.

The robot knew a priori all the objects and their locations in the experiment. No algorithms to localise or recognise objects were implemented for this experiment because the purpose was only to examine the use of different task communication methods. The robot also performed these three tasks successfully whenever requested to do so.

The speech utterances were the only interface available to the participant for communicating with the robot. The information about object locations was not communicated because, unlike in the previous experiment, both the robot and the participant knew the object names and their locations in advance.

Communication from the robot to the human was performed using Festival speech synthesis software and the robot’s manipulator movements. The main communication method was speech, through which the robot acknowledged all the participant’s task requests by describing what the robot planned to start doing next.

Two different speech-based task communication methods were defined for requesting the tasks from the robot. With the first task communication method, called the

action with object or *direct* task communication method, all of the task parameters, i.e. action and the target object, had to always be communicated explicitly in the request. With the second task communication, called the *affordance-based* or *indirect* task communication method, the task-related object name was used by itself to communicate the task. Unlike in the previous experiment, the possibility of communicating with the task-related action name was not included in order to keep the experiment duration under two hours. These direct and indirect task communication dialogues can be seen in Figure 4.12.

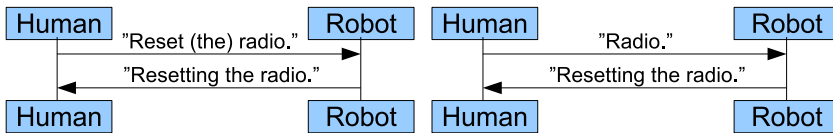


Figure 4.12: Dialogue structures of direct (left) and indirect (right) task communication methods in the second unambiguous experiment.

The direct and indirect task communications were implemented with the same dialogue manager as the previous experiment. The difference between direct and indirect task communication methods was again only in the dialogue manager and response generation parts, as were shown in Figure 4.5 and Figure 4.7.

Experimental design

The experiment used a repeated measures, i.e. within subjects, experimental design with one independent variable and three dependent variables. The independent variable was the communication method with two different levels: *direct* action with object, and *indirect* object name. These two task communication methods were presented in the previous section. The three dependent variables were the participants' task communication workload, mean task communication times, and task communication method preferences.

The experiment also included a qualitative assessment part. The goal of the qualitative assessment was to observe how participants work with the examined task communication methods, and what the participants consider to be the strengths and weaknesses of the examined system.

The experiment was counterbalanced in order to eliminate the effect of the order in which the task communication methods were used. There were two possible test round combinations because there were two levels of the independent variable, i.e. two task communication methods were tested for each of the participants. This means that only every second participant performed the experiment in exactly the same order.

Procedure

The experiment's overall mission scenario was astronaut-robot lander maintenance done on the surface of Mars. The participant was an astronaut working with the WorkPartner robot. The participant performed an inventory task while the robot merely waited for new tasks from the participant.

The experiment consisted of five different test rounds, which were performed once by each of the participants. The two first test rounds were identical except that the independent variable, i.e. the communication method, was changed for each test round. The third and fourth test rounds were identical to the first and second test rounds, respectively. The purpose of the third and fourth test rounds was to obtain experimental data from a higher point on the learning curve. In the fifth test round, both of the communication methods were available for use at the same time.

None of the test rounds was repeated by any of the participants, although the third and fourth test rounds repeated the first and second test rounds, respectively.

Further repetition of the test rounds was determined not to be necessary, since the participants repeatedly used the task communication methods during the test rounds and they were also able to rehearse using the task communications methods in advance for as long as they wanted.

The participant's goal in each of these five test rounds was the same: to fix any emerging problems by requesting the robot execute a correct task to solve the problem. The possible problems were jammed radio reception, sand built up on the solar panel, or a sample stuck in the experiment unit. To fix these problems, the participant had to request the robot either reset the radio, clean the solar panel, or pick up the measurement unit, respectively. Each of the three problems occurred two times in a random order during each of the first four test rounds and three times in the fifth test round. This means that the dialogues shown in Table 4.4 had to be repeated three times for the fifth test round and two times for each of the other four test rounds.

Table 4.4: Extract from communication dialogues of the two examined communication methods. The H→R refers to communication from human-to-robot, and R→H to communication from robot-to-human.

Description of event	Direct (action with object)	Indirect (object)
1) There is dust on the solar panel: request the robot clean the solar panel	H→R: Clean panel R→H: Cleaning the panel	H→R: Panel R→H: Cleaning the panel
2) The radio is jammed: request the robot reset the radio	H→R: Reset radio R→H: Resetting the radio	H→R: Radio R→H: Resetting the radio
3) Sample is stuck in the measurement unit: request the robot pick up the unit	H→R: Take unit R→H: Taking up the unit	H→R: Unit R→H: Taking up the unit

The flow of the experiment was as follows: The experiment scenario was first explained to participants, after which the speech recognition software was trained to correctly recognise the three object names and the three action names used in the experiment, i.e. radio, panel, unit, reset, clean and take. The participants were told to focus on solving emerging problems as quickly as possible as their primary task, and to work with a secondary inventory task, simulated by calculating arithmetic operations, only when they had free time.

The actual experiment phase started when the participants started to communicate the tasks required to solve the emerging problems in the first test round by first using only one of the two examined communication methods. Each of the three problems was shown two times in random order. The robot executed the requested tasks autonomously and always correctly, for instance, by sweeping the solar panel with a brush. The second test round was performed directly after the first test round, using the other task communication method.

Next, the two first test rounds were performed identically a second time, i.e. in the third and fourth test rounds. This time, participants filled in the NASA TLX questionnaire after each of the test rounds. The only purpose of the first and second test rounds was thus to rehearse the use of the task communication methods for the third and fourth test rounds.

Finally, in the fifth test round, the participant had to communicate tasks to the robot by freely using both of the two task communication methods at the same time. The participant was told to choose the task communication method that the participant would prefer to use for this examined mission scenario. This time, each of the three problems emerged three times. This means that the participant had to communicate a total of nine tasks to the robot in this fifth test round.

Data processing and statistical analysis

Data was collected only from the third, fourth, and fifth test rounds of the experiment. The first and second test rounds were used only to train the participants to use the task communication methods.

The participants' subjective assessments of the workload were measured using the NASA TLX, as in the previous experiment. Two NASA TLX rating sheets were collected from each of the participants. The first rating sheet was filled in by the participant right after the third test round, and the second after the fourth test round. The participants were asked to evaluate only the workload induced by task communication from the point when they noticed that a certain task had to be done to the point when they started their speech utterance. This means that, for instance, speech recognition performance was excluded from the evaluation.

The task communication times, i.e. the times from the emergence of the problems until the start of the human speech utterances, were measured for the third and fourth test rounds. These communication times were recorded during the test rounds with a stopwatch and confirmed later from video recordings. The task communication times in the third and fourth test rounds were furthermore averaged for each participant for the statistical comparison. In this case, the comparison of averages is essentially the same as in the comparison of the total task communication times, because the number of tasks was the same in both test rounds.

The number of times that the participants chose to utilise each of the compared communication methods was also counted in the fifth test round. This counting was done manually during the experiment and confirmed later from video recordings.

At the end of the experiment, the participants answered a questionnaire containing free form and multiple choice questions. The purpose of the questionnaire was to

make the participants again choose their preferred task communication method and to construct arguments for the task communication advantages and disadvantages. The qualitative part of the questionnaire was conducted as a contextual inquiry interview, similar to the first experiment.

The statistical significances of the workload, the average task communication time, and the participant's task communication choices were calculated with R software [108] using the one-way within-subjects ANOVA test. The statistical significances of the answers to the end questionnaire were calculated with a chi-square test of goodness of fit. A p value of less than 0.05 was the standard for significance.

4.2.2 Results

All of the participants eventually managed to correctly request all of the tasks that had to be accomplished in each of the test rounds. There were only a few occasions when the participant had to request a task again, for instance, due to errors in speech recognition, but eventually all of the participants managed to get the robot to execute the correct task.

The NASA-TLX subjective workload results for the direct and indirect task communication methods are shown in Figure 4.13. The one-way within-subjects ANOVA showed that the difference between the averages is significant $F(1,15)=10.29$, $p=0.006$.

The mean task communication times for the direct and indirect communication methods are shown in Figure 4.14. The one-way within-subjects ANOVA showed that the difference between the averages is significant $F(1,15)=8.027$, $p=0.013$.

The task communication method preferences of the participants, measured during the fifth test round, can be seen in Figure 4.15. The one-way within-subjects

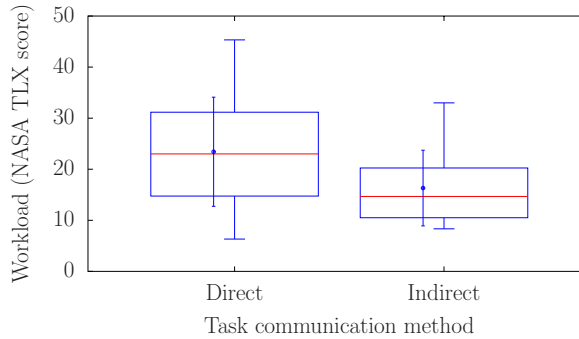


Figure 4.13: Boxplot of the NASA-TLX workloads for the compared direct and indirect task communication methods. The workload values range from 0 to 100, i.e. from no workload to full workload, respectively. Means and standard deviations are shown on the left sides of the boxplots.

ANOVA showed that the difference between the averages is significant $F(1,13)=6.650$, $p=0.023$. However, two of the 16 participants did not perform this fifth round of the experiment due to time constraints. One of these two participants started with direct task requests and the other with indirect task requests, so these results are also correctly counterbalanced.

Based on the answers to the multiple choice end questionnaire, it was found that the participants had a significant preference for indirect task communication over direct task communication, $\chi^2(1, N = 16) = 4.0, p = 0.046$.

The qualitative part of the end questionnaire provided certain insights into the potential advantages and disadvantages of affordance-based task communication. The two most frequently mentioned advantages of indirect task communication were that it was faster or easier to remember only the object name rather than both of the task parameters. These perceptions were noted by five and four participants, respectively. According to two participants, the advantage of direct task commu-

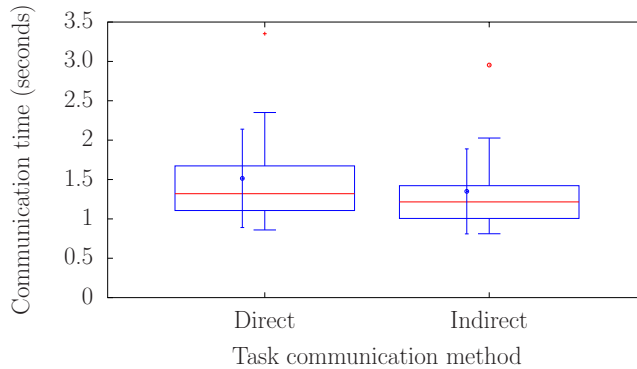


Figure 4.14: Boxplot of average task communication times for the compared direct and indirect task communication methods.

nication was that it also works without restrictions in the presence of ambiguous object-action associations.

4.2.3 Discussion

The NASA-TLX workload analysis showed that the observed workload was lower with indirect task communication than with direct task communication, as in the first unambiguous experiment. Task communication times supported this observation, as it also took in average less time for the participants to communicate by using the indirect task requests. The impact of the lower workload was probably that it was faster for the participants to formulate task requests for the robot.

The results from the fifth test round, shown in Figure 4.15, indicated that participants seemed to prefer the indirect affordance-based task requests over the direct ones. This result was congruent with the multiple-choice questionnaire results.

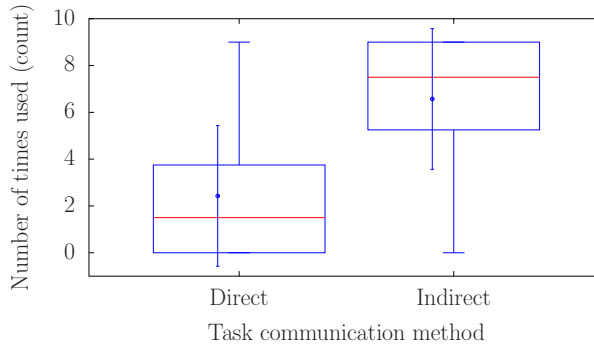


Figure 4.15: Boxplot showing utilisation of the communication methods, i.e. how many times each of the task communication methods were used by each of the participants. The maximum number of usages is nine because the total number of tasks that had to be requested was nine.

A possible explanation for these results is that with the indirect task requests the human does not need to remember the action itself, but is only required to associate which object is at the core of the task. In the case of direct communication, the human is instead required to also remember and formulate the action related to the task. The affordance-based task requests enable the human to leave object-action association as a task for the robot.

The above explanation was also posited in the first experiment. It can thus be concluded that at least in unambiguous environments, the affordance-based task communication method provides a feasible way to improve human-robot task communication with a method that humans are ready to adopt for use. The advantages can be measured both subjectively and objectively with human workload and task communication times, respectively. The next interesting question is to examine how affordance-based task communication could be extended to work in ambiguous environments where dozens of tasks are performed in the presence of ambiguous object-action associations.

5 Ambiguous Task Communication Using Affordances

The purpose of this chapter is to extend the experiments in Chapter 4 into more complex environments where each object is normally associated with several actions, and vice versa. The goal is to see if the proposed affordance-based task communication method can still be effectively incorporated into the robot.

Section 5.1 starts by presenting task request prediction methods and shows how affordances could be applicable. The overall idea is to try to utilise sequence prediction algorithms to remove ambiguities in task communication. Section 5.2 then presents a predictive dialogue approach for task communication utilising the concept of affordances. This predictive dialogue is then extended to automatic task request execution in Section 5.3.

5.1 Task request prediction

There has been a relatively long history of efforts to predict future user command sequences [23]. The overall problem addressed by these so-called Sequence Prediction Algorithm (SPA) is the determination of the conditional probability, shown in Equation 5.1, of the next input symbol x when given the sequence of i -previous input symbols $(a_1 \dots a_i)$. The input symbol x is part of the set of all possible input symbols X .

$$P(x|a_1 \dots a_i), x \in X \tag{5.1}$$

A good review of different SPAs is given by Hartmann [53]. Hartmann [53] also presents a sequence prediction algorithm called FxL, which is based on a mixed-order Markov model. The idea of the FxL algorithm is to maintain a database of different input sequence frequencies up to the desired length of k and to calculate the next symbol's probability using sequence frequencies (F) and lengths (L), as shown in Equation 5.2.

$$P(x|a_1 \dots a_i) = \frac{\sum_{j=1}^{k-1} j \cdot F(a_{i+1-j} \dots a_i \circ x)}{\sum_{y \in X} \sum_{j=1}^{k-1} j \cdot F(a_{i+1-j} \dots a_i \circ y)} \quad (5.2)$$

The FxL algorithm is chosen here as the algorithm for predicting the next task request. FxL algorithm prediction accuracy has been shown to be between 43% and 58%, with a 90% applicability level, when predicting different computer programs' user commands [53]. Prediction accuracy is defined as a ratio between the number of correct predictions that were over the probability threshold used and the number of all predictions that were over the probability threshold used. Applicability is the ratio between the number of times when there was one or more predictions over the required probability threshold and the number of times when there was one or more points of history available to make a prediction.

For example, an applicability of 90% thus means that 90% of the time the algorithm is able to give a prediction that has a higher probability than the probability threshold used. For example, a 43% prediction accuracy means that 43% of these predictions with a probability over the probability threshold used were correct ones. An increase in the probability threshold generally causes applicability to decrease, but the prediction accuracy to increase.

5.1.1 Prediction with affordances

One part of the approach examined here is to use the human communicated object or action name to further restrict the predicted task request. This means that after we have listed the most likely next tasks using a sequence prediction algorithm, which is FxL in this case, we further limit this list of the most likely tasks by considering only tasks that include the communicated object or action name. For example, if the human communicates “rock” and our most likely tasks are “analyse rock” and “pick up unit”, then we would consider only the “analyse rock” task because it is the only task containing the “rock” object. In addition, as a final option, if no usable predictions were found, completion of the task request is attempted using the action or object name from the previous task.

An indication of potential task prediction accuracy when using FxL and affordance-based task requests can be obtained by evaluating the algorithm using existing datasets. Figure 5.1 shows evaluation results that were obtained for this thesis with the affordance-based method using a dataset containing logs of Microsoft Word usage [80]. The affordance-based communication was simulated by extracting either the object or action part of each task request, respectively. For example, if the correct request is “FileOpen” then the user would communicate the “File” object or the “Open” action using the affordance-based method.

Figure 5.1 clearly shows that the use of affordances can significantly increase prediction accuracy. Prediction accuracy stabilises at around 70%, while incorrect predictions comprise around 20% of the predictions. For the last 10% of the predictions, there are no usable predictions given by the algorithm. Use of an action hint instead of an object hint was able to give a slightly better prediction accuracy, at least for this dataset. The reason is probably that different actions are often performed with certain object, rather than carrying out the same action with different objects.

The Word dataset [80] was interpreted so that each unique “user”-“file size” pair is a new usage session. The usage sessions are then ordered according to starting time and run through the algorithm. Finally, the results are macro averaged, i.e. the average is calculated for all users independent of the length of their dataset.

5.2 Predictive dialogue experiment

This third user experiment extends the scope of the first two user experiments, presented in Chapter 4, to situations containing dozens of different tasks and ambiguous object-action associations. The increased complexity required the experiment to be implemented with a video-based robot simulator. This time the overall context of the experiment was an astronaut-robot lander assembly mission.

5.2.1 Method

Participants

A total of 18 participants were selected for the experiment. Three of the participants were female and 15 were male. The average age of the participants was 26.7 ± 5.5 years. All of the participants were either Aalto or Helsinki University students or researchers. Two of the participants were native English speakers. All of the participants can be considered novices because they did not have any previous experience with the examined system. Participants were compensated for their participation with a movie ticket.

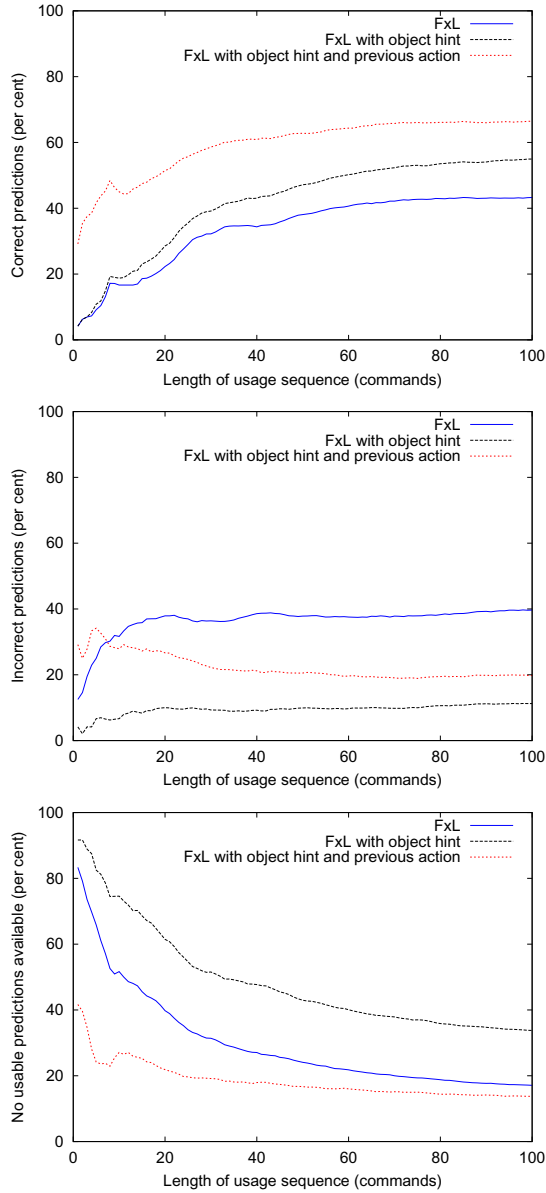


Figure 5.1: Correct (top), incorrect (middle) and non-possible (bottom) predictions with a prediction applicability of one, and when using the extracted object name as a communication utterance hint.

Equipment and software

The physical setup of the experiment is shown in Figure 5.2. As in the first user experiment, the participants had a Shure PG1 wireless microphone attached to their chest for speech communication. The participants, who sat in the black chair for the duration of the experiment, were also given a sheet of paper full of uncompleted multiplications. This kind of multiplication task is a commonly used secondary task in user studies [44].

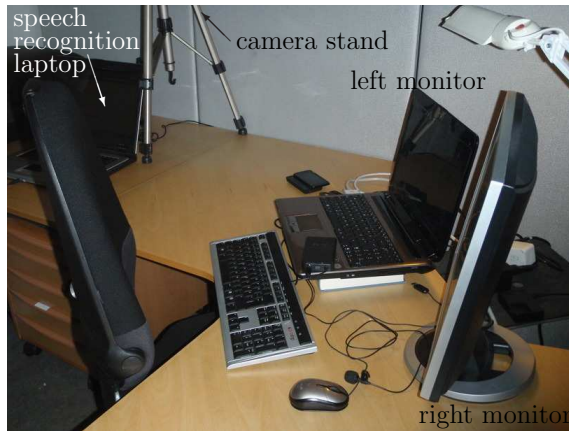


Figure 5.2: Experiment setup for the ambiguous task communication experiment. The laptop in front of the chair was used to run the simulator (left monitor) and to show a picture depicting the next task to be requested (right monitor). The laptop at the back (behind the chair) was used to run the speech recognition software.

The software architecture used in the experiment is described in Section B.3. The speech recognition software used in the experiment was the commercial Nuance Dragon NaturallySpeaking 10.0³. The speech recognition software output was processed, as in the first experiments, with the frame-based dialogue manager shown in Figure 4.5 and Figure 4.7. For this experiment, the dialogue manager was modified to enable it to solve ambiguous object-action associations with an affordance-based

³<http://www.nuance.com/dragon/>

dialogue, in case querying the object-action database returned several possible tasks. The participant was, however, able to request tasks in two different ways: (i) with a direct question-based dialogue, and (ii) using an indirect affordance-based dialogue.

The direct question-based dialogue is the current conventional solution to ambiguous task communication, as argued in the review presented in Chapter 2. The idea is that the robot can reply with a list of all the objects it knows or the actions that it can perform with a certain object [66, 77]. This also uses the concept of affordances at a certain level, because the robot replies are formulated using known object-action associations. It also already enables the astronaut to communicate any tasks that might be required. Significant disadvantages of this kind of mechanical listing are the long time required to do the listing and the unnecessarily high workload caused by the listing.

The indirect affordance-based task communication method was formulated for this experiment based on the experiences gained from the unambiguous task communication experiments. The hypothesis was that the object or action names alone could be used to communicate the tasks more efficiently. The object-action ambiguities were resolved by using past task requests to predict the most likely next task requests. These predictions made by the robot were then accepted or rejected by the participant.

The algorithm used for predicting the requests from the task history was a mixed-order Markov model-based FxL sequence prediction algorithm [53], which was described with detail in Section 5.1. The sequence of tasks in the experiment was fine-tuned so that 75% of the affordance-based task request predictions were correct with the FxL algorithm. This prediction rate was selected based on FxL algorithm performance with human-computer interaction data, such as Microsoft Word usage [80]. The underlying assumption was thus that the tasks are often performed in predictable sequences that can be learned while a mission is being performed. The

Algorithm 1 Dialogue manager pseudo-code that was used to implement the affordance-based task communication methods in the ambiguous user experiments.

```

newTaskRequest = FormFillingMethod(utterances)
if newTaskRequest == PARTIAL_REQUEST then
    predictedTasks = TaskRequestPredictionWithFxl(newTaskRequest)
    predictedTasks = TaskRequestPredictionWithPreviousTask(previousTask)
end if
executableTask = RequestedTaskIsAllowed(newTaskRequest, predictedTasks)
if executableTask == ALLOWED then
    ExecuteTaskRequest(executableTask)
end if

```

dialogue manager pseudo-code that is able to solve affordance-based task communication ambiguities is shown in Algorithm 1.

The participants communicated with a simulated WorkPartner robot in the experiment, because a complex experiment like this would have been very difficult to control and implement with a real robot. The simulated WorkPartner system was identical to the real WorkPartner robot, except that the task execution modules were replaced with a module playing video sequences, as shown in Figure 5.3. These *a priori* recorded video sequences showed the real WorkPartner robot performing the requested tasks. In total there were 21 possible actions and 6 target objects, which enabled WorkPartner to perform 65 different tasks when counting only the possible object-action associations. These 65 tasks are listed in Table 5.1.

OpenOffice.org Impress⁴ was used to display the mission task sequence to the participants. The participants were able to see a picture depicting the next task to be performed by pressing any key on the keyboard after previous task execution had finished. Twenty such task description pictures are shown in Figure 5.4.

⁴<http://www.openoffice.org/product/impress.html>

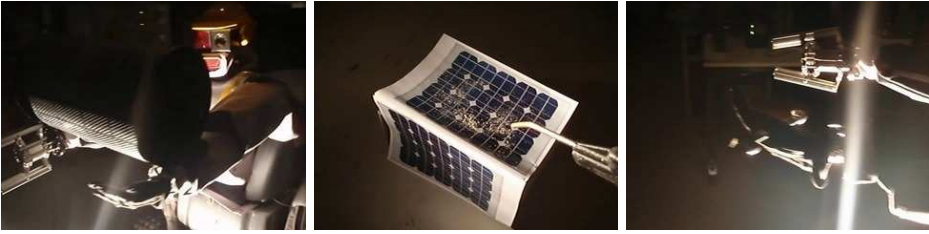


Figure 5.3: Screen shots from the video-based robot simulator that was used to visualise the robot’s task execution for the participants. The screen shots show WorkPartner holding a battery pack (left), cleaning a solar panel (middle), and handing someone a wrench (right).

In this experiment, the participant was able to communicate with the robot only with speech. However, the participant had two different types of speech-based task communication methods available. The first was the so-called direct task communication method where, as in the first experiment, an action with object utterance was always used as the final utterance to request the task. The second was the so-called indirect task communication method where object or action names were used by themselves to request a task. The dialogue structures of the direct and indirect task communication methods are shown in Figure 5.5.

The direct task communication method presented the current conventional solution to ambiguous task communication, which is based on the robot’s ability to list the objects it knows and the actions that it can perform. These lists helped participants to remember the task request utterance by reminding them of the action and object names related to the task. The object and action listings were always presented in the same order in which they were originally randomly set for the experiment.

The indirect task communication utilised the concept of affordances, as in the unambiguous experiments, by enabling only the object or action name to act as a task request. The robot predicted the most likely task request if the object or action name did not unambiguously define a task. The human had to accept or reject the

Table 5.1: List of all possible tasks, i.e. actions with different objects, in the ambiguous experiment. The six objects in the experiment were wrench, NASA module, JAXA module, solar panel, battery pack, and radio antenna.

Task	Object(s)	Action	Description of task
1-6	All six	pick up	Takes the object from a current location
7-12	All six	insert	Places the object in a given location
13-18	All six	image	Takes a picture of the object
19-24	All six	store	Takes the object to a storage place
25-29	All but wrench	forward	Moves the object forward
30-34	All but wrench	backward	Moves the object backward
35-39	All but wrench	hold	Holds the object in the current location
40-44	All but wrench	rotate	Enables the human to rotate the object
45-49	All but wrench	power on	Connects the object to a power bus
50-54	All but wrench	power off	Disconnects the object from a power bus
55	wrench	bring	Moves the wrench close to the requester
56	battery pack	measure	Measures the voltage of the battery pack
57	solar panel	clean	Removes dust from the solar panel
58	JAXA module	analyse	Analyses the condition of the module
59	JAXA module	reboot	Does a software reset for the module
60	JAXA module	reset	Does a hardware reset for the module
61	NASA module	calibrate	Calibrates the module
62	NASA module	shake	Shakes the module to spread the sample inside
63	radio antenna	erect	Erects the antenna for use
64	radio antenna	tune	Finds the optimal frequency for transmission
65	radio antenna	point	Finds the optimal pointing direction

robot’s task prediction by replying either “yes” or “no”, or alternatively by correcting the task request with the right object or action name. As a final option, the robot listed all the associated object or action names if the participant replied “no” twice to the robot’s task request predictions.

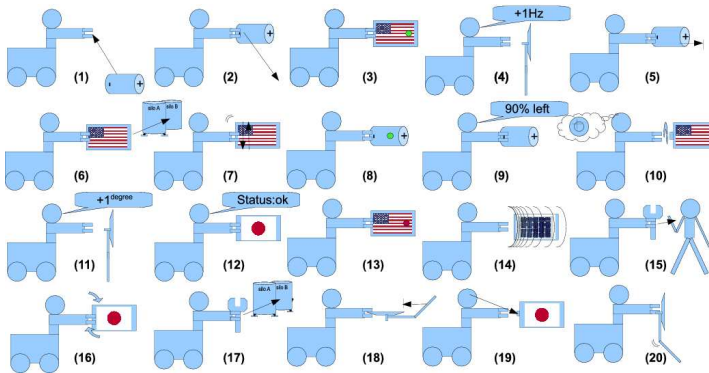


Figure 5.4: Twenty pictures depicting tasks performed by the participants in the first two test rounds of the experiment. The first task, for example, instructs the participant to request a “pick up battery pack” task.

Communication from the simulated robot to the human was performed through Festival speech synthesis software, and the robot’s location and orientation in the video. Speech was the main communication method through which the robot acknowledged all the participant’s task requests and requested confirmation of task request predictions.

Experimental design

The experiment used a repeated measures experimental design with one independent variable and three dependent variables. The independent variable was a task communication method with two different levels: direct and indirect. These direct and indirect task communication methods were presented in the previous section. The three dependent variables were the participants’ task communication workload, the total test round execution time, and the participants’ task communication preferences.

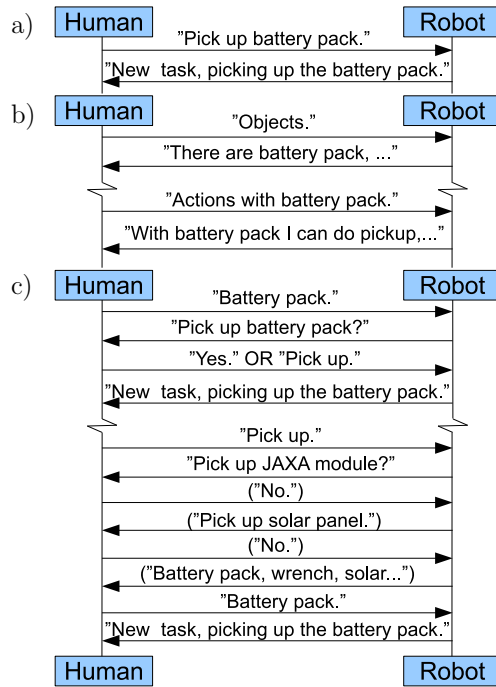


Figure 5.5: The dialogue structures of the direct (b) and indirect (c) task communication methods. In the shared dialogue (a) the tasks are requested using explicit action with object task utterances.

The experiment also included a qualitative assessment part. The goal of the qualitative assessment was, as in the previous experiments, to observe how the participants work with the examined task communication methods, and what they considered to be the strengths and weaknesses of the examined system.

The experiment was counterbalanced in order to eliminate the effect of the order in which the task communication methods were used. The number of counterbalanced test rounds was two because there were two levels of the independent variable, i.e. two task communication methods were tested for each of the participants. This means that only every second participant performed the experiment in exactly the same order.

Procedure

The overall scenario in the experiment was an astronaut-robot lander preparation on Mars. The experiment consisted of three different test rounds, which were performed once by each of the participants. The first two test rounds were identical except that the independent variable, i.e. the task communication method, was changed for each round. In the third test round, both the direct and indirect task communication methods were available for use at the same time. None of the test rounds was repeated by any of the participants. It was not deemed necessary to repeat the test rounds, since the participants repeatedly used the task communication methods during the test rounds and they were also able to rehearse using the task communications methods in advance for as long as they wanted.

The goal of each of these three test rounds was the same: to communicate 20 tasks - displayed one by one on the monitor - like an astronaut would do when working on Mars. Figure 5.4 shows the 20 tasks communicated in each of the first two test rounds. Another different set of 20 tasks was communicated in the third round.

The flow of the experiment was as follows: After hearing a description of the experiment's overall mission scenario, the participant was informed about all the objects and actions available in the experiment. Each task, consisting of an action performed on a certain object, was described to the participant with a comic strip type of picture, as shown in Figure 5.4. After learning to recognise all the tasks from these pictures, the participant trained the speech recognition software to correctly recognise all the words used in the experiment.

Next, after receiving an explanation of how the compared communication method dialogues worked, the participants tried all of the possible dialogue options shown in Figure 5.5 a few times. Depending on the participant, this required five to ten rehearsal task communications.

The actual experiment started when the participant was instructed to start to communicate the 20 tasks of the first test round. An example of task communication with both direct and indirect task communication methods is shown in Table 5.2.

Between task request communications, while the robot executed the requested task, the participant calculated multiplications given on a sheet of paper as a secondary task. After successfully completing a test round, the participant filled in a NASA TLX questionnaire before starting the next test round.

All three test rounds were completed one right after the other. Participants were told after each test round which communication method was to be used next. In the third test round, participants were instructed to choose the communication method that they would prefer to use if they were astronauts working on Mars. The introduction and rehearsal phase of the experiment took approximately 50 minutes, and the three test rounds and the final questionnaire took around 40 minutes.

Table 5.2: Example of typical communication dialogue between the Human(H) and the Robot(R) in the ambiguous experiment.

Event description	Direct method	Indirect method
1) Requesting the robot to pick up the solar panel using action and object	H→R: Pick up the solar panel R→H: New task, picking up the solar panel	H→R: Pick up the solar panel R→H: New task, picking up the solar panel
2) Requesting the robot to take an image of the wrench, but without knowing the object's name	H→R: Objects R→H: There are wrench, battery... H→R: Image wrench R→H: New task, taking image of the wrench	H→R: Image R→H: Image wrench H→R: Yes R→H: New task, taking image of the wrench

Data processing and statistical analysis

The participants' subjective assessments of the workload were measured using NASA TLX [52]. Two NASA TLX rating sheets were collected from each of the participants, one after the direct task communication test round and another after the indirect task communication test round. The participants were asked to evaluate only the workload induced by the task communication, from the point when they understood what task had to be requested to the point when they were sure that the robot was executing the correct task.

The total time to complete a test round was measured from the first human task request utterance to the last human task request utterance. These test round completion times were extracted from the dialogue manager's log files.

The participants' task communication preferences were collected from the third test round by counting the number of times that the participants chose to utilise each of the two compared communication methods. This counting was done manually during the experiment and confirmed later from video recordings and from the dialogue manager's log files.

At the end of the experiment, the participants answered a questionnaire containing free form and multiple choice questions. The purpose of the questionnaire was to make the participants again choose their preferred task communication method and to construct arguments for the task communication advantages and disadvantages. The qualitative part of the questionnaire was conducted as a contextual inquiry interview, similar to the first experiments.

The statistical significances of the workload, the round completion time, and the participant's task communication choice results were calculated with R software [108] using the one-way within-subjects ANOVA test. The statistical significances

of the answers to the end questionnaire were calculated with a chi-square test of goodness of fit. A p value of less than 0.05 was the standard for significance.

5.2.2 Results

All of the participants managed to request correctly, and in the right order, all 20 tasks required to accomplish the lander assembly mission in each of the test rounds. Some of the participants occasionally had to request a task again, for instance, due to the use of incorrect words or errors in the speech recognition, but eventually all of them managed always to get the robot to execute the correct task.

The NASA-TLX subjective workload results for the direct and indirect task communication methods are shown in Figure 5.6. The one-way within-subjects ANOVA showed that the difference between the averages is significant $F(1,17)=11.70$, $p=0.003$.

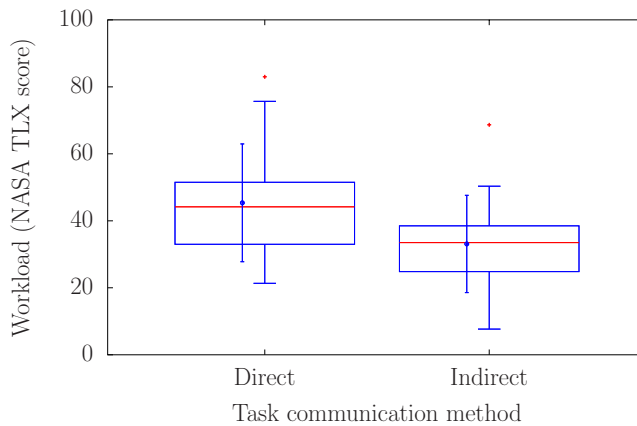


Figure 5.6: Boxplot of the NASA-TLX workloads for the compared direct and indirect task communication methods. Workload values range from 0 to 100, i.e. from no workload to full workload, respectively. Means and standard deviations are shown on the left sides of the boxplots.

The total test round execution times while using the indirect and direct communication methods are shown in Figure 5.7. The one-way within-subjects ANOVA showed that the difference between the averages is significant $F(1,17)=11.27$, $p=0.004$.

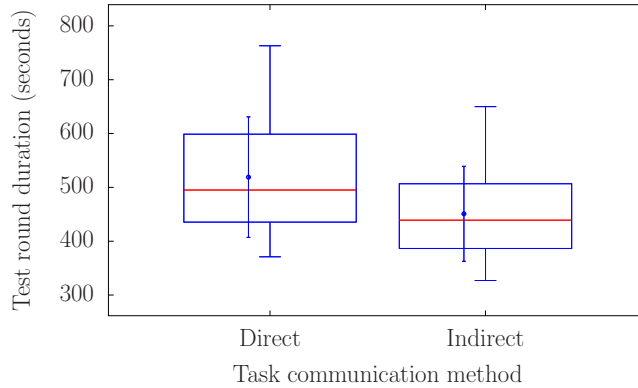


Figure 5.7: Boxplot of test round execution times for the compared direct and indirect task communication methods.

The participants' communication method preferences, measured in the third test round, can be seen in Figure 5.8. The one-way within-subjects ANOVA showed that the difference between the averages is significant $F(1,17)=7.94$, $p=0.012$.

Based on the answers to the multiple-choice end questionnaire, it was found that participants had a significant preference for indirect task communication over direct task communication, $\chi^2(1, N = 18) = 8.0, p = 0.0047$. The participants were also found to prefer using both direct and indirect task communication at the same time over having only either direct or indirect task communication available, $\chi^2(1, N = 18) = 14.2, p = 0.0002$.

The qualitative part of the end questionnaire provided some insights into the potential advantages and disadvantages of affordance-based task communication. The two most frequently mentioned advantages of indirect task communication were that

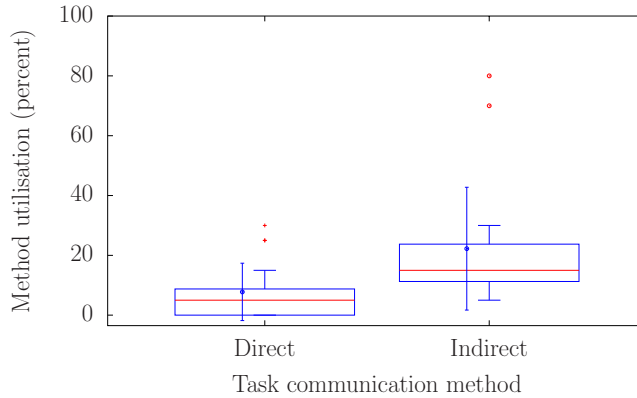


Figure 5.8: Boxplot of percentages of use of communication methods, i.e. the portion of the task request in which either the direct or indirect communication method was used, for the third round of the experiment. The rest of the task requests were explicit task requests containing both action and object names.

it does not require any additional syntax, as object and action names are already known, and that it is easier to remember only an object or action name than both of them.

The two most frequently mentioned advantages of the direct task communication method were its ability to also work when both the object and action names are unknown, and its dialogue performance that does not depend on the robot’s task request predictions. These advantages of the direct task communication method are equally disadvantages of the indirect task communication method, and vice versa.

In the end, the number of task requests, where the participants used something other than the shared explicit action with object utterances, were also counted for all of the three test rounds. These counts are shown in Figure 5.9.

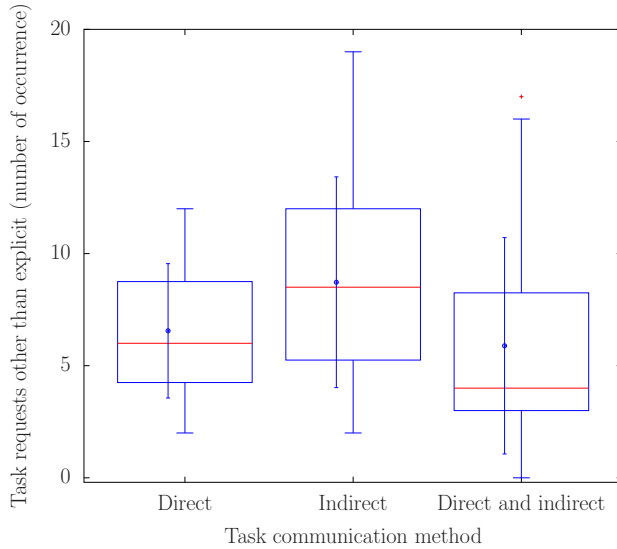


Figure 5.9: Boxplot showing the number of task requests in each of the three performed test rounds where something other than explicit action with object task utterances were used. The maximum value is 20 because there were 20 task requests in each of the test rounds.

5.2.3 Discussion

The main finding of the experiment was that the formulated indirect task communication method was able to simultaneously decrease the subjective human workload and the total test round execution times, while also being the preferred way to communicate tasks. This is a clear indication that the proposed affordance-based indirect task communication method is a feasible and effective way to improve explicit speech-based human-robot task communication in complex work environments as well.

This result is congruent with the unambiguous experiments as the affordance-based task communication methods were shown to decrease the participants' subjective workload in all cases. The main argument in favour of indirect task communication

was also the same in the experiments, i.e. it is easier to remember only a task's object or action name than both of the names. There did not seem to be any significant additional mental processing, such as thinking about the object-action associations, that would have hindered the task communication.

All except one of the participants answered that they would prefer that both the direct and indirect task communication methods could be used at the same time, although in general they preferred to use indirect task communication. This is not a surprising result, because the proposed indirect task communication method by itself is not sufficient in all possible situations, as for example, the human does not know what objects the robot knows and what actions it can perform with the objects.

In the first two test rounds, all of the participants chose to utilise something other than explicit utterances, i.e. other than action with target object utterances, when requesting tasks, as shown in Figure 5.9. This is congruent with the well-known finding that humans cannot be expected to remember several fixed communication utterances [42, 73, 65].

The experiment was performed with the assumption that the robot is able to correctly predict more than 75% of the requested tasks. In other words, this means that 75% of the task requests must be part of already performed or otherwise known task sequences, because predictions are made based on the task request sequences known by the robot. This is not an unreasonable requirement, especially for the examined planetary exploration missions where the performed tasks are usually carefully planned well in advance. Nonetheless, the task communication remains usable even if only unexpected tasks are performed, because the affordance-based indirect task communication method does not replace any existing functionality but can be instead used as a supplementary task communication method.

Another assumption in the experiment was that the tasks consist of actions and target objects. The context of the experiment, i.e. an astronaut-robot lander assembly mission, had components similar to the facility maintenance scenario described in Chapter 3. That mission, like all the other missions in Chapter 3, were all constructed using tasks containing actions and target objects. This indicates that there should not be any real constraints to extend the use to other types of missions based on the selected experiment. The external human-computer interaction data, which was used to select task prediction accuracy, also had around ten times more objects and two times more actions than the performed experiment. Given this, communication can likewise be expected to scale up to more complex scenarios with a similar performance.

The formulated indirect task communication method was only one possible way to implement the affordance-based task communication. One of the next questions is whether the indirect task communication could be further improved by eliminating the task request confirmation dialogues. The idea is to make task communication easier for the most likely case when the prediction is correct, and require the human to communicate further only if the prediction was not correct. The next experiment continues to explore the potential of affordance-based task communication based on this idea.

5.3 Automatic execution experiment

The purpose of this experiment is to further analyse task communication methods inspired by the concept of affordances in ambiguous work environments. This experiment specifically examines whether the affordance-based task communication could be further improved by removing all the task request confirmation dialogues. This experiment's overall mission scenario is the same as in the previous experiment, i.e. cooperative astronaut-robot lander assembly.

5.3.1 Method

Participants

A total of 16 participants were selected for the experiment. Fourteen of the participants were male and two were female. The average age of the participants was 24.6 ± 4.4 years. Thirteen of the participants were Aalto, Helsinki or Oulu University students, while three were working in companies. None of the participants spoke English as their native language. All of the participants can be considered novices, as they did not have any previous experience with the examined system. Participants were compensated for their participation with a movie ticket.

Equipment and software

The experiment setup was almost identical to the previous experiment, described in Section 5.2.1. The only two changes were the communicated tasks and the compared task communication methods itself.

The number of communicated tasks in this experiment was 40, as shown in Figure 5.10. Thirty of the tasks were communicated according to a predefined nominal task sequence. The ten remaining tasks occurred without *a priori* knowledge on the part of the robot or the participant. The robot's task request prediction algorithm was initialised using the sequence of 30 nominal tasks. For this reason, the robot was always able to predict a nominal task correctly if the previous task had also occurred in the nominal task sequence just before that task. Together with the 10 unexpected tasks, this means that 17 out of the 40 tasks were not predicted correctly.

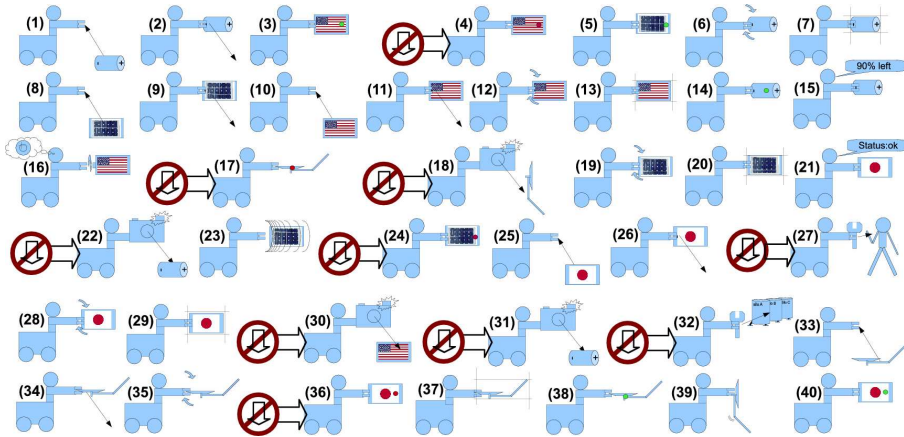


Figure 5.10: List of 40 tasks performed by the participants. The ten tasks with red prohibition signs were tasks that the robot could not predict correctly when using the indirect task communication method. The other 30 tasks were performed according to the expected nominal task sequence.

The two task communication methods used to communicate tasks were the *dialogue*-based and *automatic execution*-based methods. The *dialogue*-based task communication method always tried to solve the task communication ambiguities by initiating a dialogue, while the *automatic execution*-based method executed the task automatically when the task was calculated to be probable enough.

Both these task communication methods were considered here to be affordance-based task communication methods, because in both of the methods the task request can be initiated using only the action or object name related to the requested task. The dialogue structures of these task communication methods are shown in Figure 5.11.

The idea of the *dialogue*-based task communication method was to allow the participant to initiate a task request with only the object or action name in case the participant was not able to remember the full task request utterance. In that case, the robot gave a full list of all the associated actions or objects in an order sorted

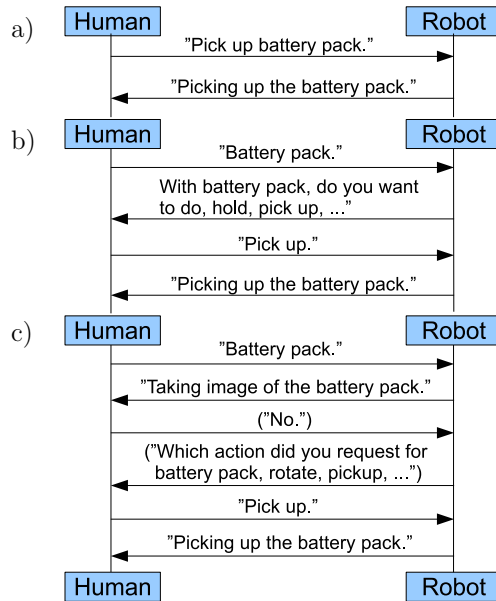


Figure 5.11: The dialogue structures of the *dialogue*-based (b) and *automatic execution*-based (c) task communication methods. In the shared dialogue structure (a) the tasks are requested using explicit action with object task utterances.

according to predictions of most likely task requests. However, the human always had to complete the task request by stating the missing object or action name.

The *automatic execution*-based task communication method instead tried to minimise the amount of required dialogue by removing the need to confirm the task request to the robot. This means that the participant had to request the task again if the robot did not start to execute the correct task. The new request could be the right object or action name, or a “no” utterance. With the right object or action name, the robot switched directly to execute the correct task; with the “no” utterance, the robot gave instead a list of all possible associated objects or actions. The robot stated immediately which task will be executed, but the actual task execution did not start until one second later. The only exception was the “stop” task request, which was executed immediately.

Experimental design

The experiment used a repeated measures experimental design with one independent variable and three dependent variables. The independent variable was a communication method with two different levels: *dialogue*-based and *automatic execution*-based task communication methods. These two communication methods were presented in the previous section. The three dependent variables were the participants' task communication workload, the test round execution time, and the number of times affordance-based task communication was used.

The experiment also included a qualitative assessment part. The goal of the qualitative assessment was, as in the previous experiments, to observe how the participants work with the examined task communication methods, and what they considered to be the strengths and weaknesses of the examined system.

The experiment was counterbalanced in order to eliminate the effect of the order in which the task communication methods were used. The number of possible test round combinations was two because there were two levels of the independent variable, i.e. two task communication methods were tested for each of the participants. This means that every second participant performed the experiment in exactly the same order.

Procedure

The overall scenario in this experiment was astronaut-robot lander preparation on Mars, as in the previous experiment. The participant, who was acting as an astronaut, had to use speech utterances to request 40 tasks from the robot in order to successfully complete each of the experiment's test rounds.

The experiment consisted of two different test rounds, which were both performed once by each of the participants. The two test rounds were identical except that the independent variable, i.e. the task communication method, was changed for each round.

The test rounds were not repeated by any of the participants. It was not deemed necessary to repeat the test rounds, since the participants repeatedly used the task communication methods during the test rounds and they were also able to rehearse using the task communications methods in advance for as long as they wanted.

The goal of both of the test rounds was the same: to communicate 40 tasks - shown one by one on the monitor - as an astronaut would do when working on Mars. Figure 5.10 shows the 40 tasks communicated in both of the test rounds.

The flow of the experiment, which is very similar to the previous experiment, is as follows: After hearing a description of the experiment's overall mission scenario, the participant was informed about all the objects and actions available in the experiment. Each task, consisting of an action performed on a certain object, was described to the participant with a comic strip type of picture, as shown in Figure 5.10. After learning to recognise all the tasks from these pictures, the participant trained the speech recognition software to correctly recognise all the words used in the experiment dialogues.

Next, after receiving an explanation of how the compared communication method dialogues worked, the participants tried all of the possible dialogue options once. Depending on the participant, this required three to six rehearsal task communications.

The actual experiment phase started when the participant was instructed to start to communicate the 40 tasks of the first test round. An example of task communication with the *dialogue*-based and the *automatic execution*-based task communication

methods is shown in Table 5.3. Between task communication requests, while the robot executed the requested task, the participant calculated multiplications given on a sheet of paper as a secondary task, just as in the previous experiment.

The two test rounds were completed one immediately after the other. The participant was told after each test round which communication method was to be used next. The introduction and rehearsal phase of the experiment took approximately 50 minutes, and the two test rounds and the final questionnaires took around 40 minutes.

Data processing and statistical analysis

The participants' subjective assessments of the workload were measured using the NASA TLX [52]. Two NASA TLX rating sheets were collected from each of the participants, one after each of the test rounds. The participants were asked to evaluate only the workload induced by the task communication, from the point when they understood what task had to be requested to the point when they were sure that the robot was executing the correct task.

The total time to complete a test round was measured from the first human task request utterance until the last human task request utterance. These test round completion times were extracted from the dialogue manager's log files. At the end of the experiment, the participants also filled in a questionnaire containing free form and multiple choice questions.

The number of times affordance-based task communication was used was counted for the two test rounds. The task requests were classified as affordance-based if there was more than a one-second pause between the action and the target object utterances.

Table 5.3: Extracts from the participants’ actual communication dialogues for the two examined communication methods. The H→R refers to communication from human-to-robot and R→H to communication from robot-to-human.

Description of the event	Dialogue-based	Automatic execution-based
1) Request the robot to pick up a battery pack using direct task communication.	H→R: Pick up battery pack R→H: Picking up the battery pack	H→R: Pick up battery pack R→H: Picking up the battery pack
2) Request the robot to insert a battery pack using indirect affordance-based task communication. Prediction is correct.	H→R: Battery pack R→H: With battery pack, do you want to do insert, rotate, ... H→R: Insert R→H: Inserting the battery pack	H→R: Battery pack R→H: Inserting the battery pack
3) Request the robot to store a solar panel using indirect affordance-based task communication. The task prediction fails.	H→R: Store R→H: Do you want to store a wrench, solar... H→R: Solar panel R→H: Storing the solar panel	H→R: Store R→H: Storing the wrench H→R: Solar panel R→H: Storing the solar panel
4) Request the robot to rotate a radio antenna using indirect affordance-based task communication. The task prediction fails and participant does not remember the task-associated action name.	H→R: Radio antenna R→H: With radio antenna, do you want to do insert, rotate, ... H→R: Rotate R→H: Rotating the radio antenna	H→R: Radio antenna R→H: Inserting the radio antenna H→R: No R→H: With radio antenna, do you want to do insert, rotate, ... H→R: Rotate R→H: Rotating the radio antenna

At the end of the experiment, the participants answered a questionnaire containing free form questions. This qualitative questionnaire was conducted as a contextual inquiry interview, as in the other experiments.

The statistical significances of the workload, the round execution times, and the results for the number of times the affordance-based task communication was used were calculated with R software using the one-way within-subjects ANOVA test. A p value of less than 0.05 was the standard for significance.

5.3.2 Results

All of the participants managed to request correctly, and in the right order, all the 40 tasks required to accomplish the lander assembly mission in each of the test rounds. The degree of use of the affordance-based task communication method is shown with a boxplot diagram in Figure 5.12. The one-way within-subjects ANOVA showed that the difference between the averages is significant $F(1,15)=8.27$, $p=0.012$.

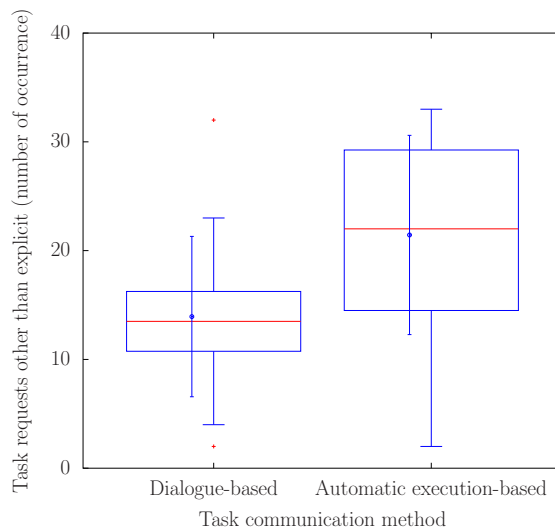


Figure 5.12: Boxplot showing the number of task requests where something other than explicit action with object task utterances were used. The means and standard deviations are shown on the left sides of the boxplots. The maximum value is 40 because there were 40 task requests in each of the test rounds.

The recorded task communication times and the collected NASA TLX data did not show statistically significant differences between the task communication methods. The advantages and disadvantages that were mentioned most often by the participants in the free form questionnaire are listed in Table 5.4.

Table 5.4: Advantages and disadvantages of the examined task communication methods based on the free form questionnaire. The number of participants arguing for that specific point is given in parentheses.

Communication method	Advantages	Disadvantages
<i>Dialogue</i> -based	Good predictability (6)	More repetitive dialogue feels boring (6)
<i>Automatic execution</i> -based	Practical for doing different actions with the same object (5)	Incorrect predictions feel risky (4)

5.3.3 Discussion

The main finding of this experiment was that the two examined affordance-based task communication methods were also successfully used to communicate tasks to the robot. The *automatic execution*-based communication method can be used to create more naturally flowing dialogue, but at the cost of executing potentially dangerous tasks.

This trade-off between task communication flow and risk was visible in the way that some participants always preferred to play it safe and utilise predictions only as their last possible option, while others found it easier to accept predictions and deal with the correction dialogue in case the prediction was wrong. In a planetary exploration context, the *automatic execution*-based task communication method would probably not be a very viable option because risk minimisation is a very high priority.

The importance of risk minimisation, for instance, could have been taken into account with additional metrics. However, it can be argued that additional metrics, such as risk-related trust, are partly included in the other metrics that were used. For instance, automatic task execution seemed to increase the human workload and make the participants prefer the other compared task communication method; it was argued that this was due to a lack of trust, as shown by the free-form questionnaire. Nonetheless, additional metrics might have provided quantitative results to support these qualitative explanations about differences in performance and preferences.

The degree of use of the affordance-based task communication method was congruent, as in the previous experiment, with the finding that humans cannot be expected to remember several fixed communication utterances. All the participants in this experiment experienced situations where they could not remember one of the two task communication parameters. However, all the participants were able to remember at least one of these parameters, because all the participants were able to communicate all 40 tasks successfully.

Indirect affordance-based task communication was utilised more with the *automatic execution*-based task communication method than with the *dialogue*-based task communication method. This can be explained by the difference in potential utility provided by the task communication methods. The *automatic execution*-based task communication method was able to assist the participant by immediately executing the most likely task request, while the *dialogue*-based task communication method was useful in practise only when the participant was not able to remember the action or the object name.

6 Conclusion

This thesis formulated a new affordance-based task communication method for the purpose of face-to-face astronaut-robot task communication. The idea of the affordance-based task communication method is to give the robot a human-like ability to understand affordances, i.e. action possibilities, in task communication. With the affordance-based method, astronauts are able to communicate tasks using only the task-related action or target object names, and thus avoid the need to remember full task request utterances.

Four user experiments were performed to analyse the usefulness of affordance-based task communication. The first two user experiments, performed with a fully autonomous WorkPartner robot, indicated that humans are capable of, and willing to, communicate tasks with the affordance-based task communication methods, and that the user task workload can be reduced in comparison with conventional task communication methods. Furthermore, the second user experiment also showed that task communication times can be decreased.

The third user experiment extended the first two user experiments from unambiguous work environments, where each action is associated with one object, to ambiguous environments, where several actions are usually associated with each object, and vice versa. The ambiguities of the task requests were solved by predicting the next task based on past task request sequences and by using a speech-dialogue to confirm the predicted task requests. The results again showed a decrease in task communication workloads and task communication times. The fourth user experiment indicated that automatic execution of ambiguous task requests is not very usable for the astronaut-robot planetary exploration work context owing to the elevated risk of executing potentially dangerous tasks, even though the affordance-based task communication dialogue might therefore be more fluent.

The affordance-based task communication methods formulated also resemble speech-based menus, which have been used in the past to communicate with both computers and robots. However, the novelty of the presented affordance-based task communication method lies in structuring the menus to use object-action associations. Graphical user interfaces have used context-menus that operate similarly by displaying actions related to a selected object, but not usually the other way round, as is done here by displaying objects that are related to a certain action.

From the human perspective, affordance-based task communication methods do not necessarily even appear menu-like, because the robot communicates with natural language sentences. The affordance-based task communication dialogues were formulated in this thesis so that task communication resembles human-human discussion about the requested task rather than appearing as a speech-based menu.

The ability of robots to interpret object or action names as task requests was the common factor in all of the affordance-based task communication methods presented in this thesis. The experiments showed that humans find it logical to request tasks through object and action names. The observed decrease in the human task communication workload and in task communication times can be explained by the fact that the affordance-based task communication methods did not introduce any additional syntax, and that they allow for opening the task dialogue by remembering only one object or action name.

In this thesis, affordance-based task communication was implemented along with a frame-based dialogue manager. This means that affordance-based task communication could be readily integrated into many existing robots, because frame-based dialogue managers are well known and widely used in robotics. Other types of dialogue managers should also be usable, since the only requirement from the dialogue manager is the ability to keep track of the task request history and possible object-action associations.

The two user experiments with ambiguous action-object associations incorporated the additional assumption that more than 75% of the requested tasks had to be *a priori* known or already performed task sequences. This requirement is acceptable, for instance, for the examined astronaut-robot planetary exploration target environment, because the task types to be performed are usually carefully planned in advance.

The applicable use scenarios of affordance-based task communication should not, however, be considered limited to the examined robotic astronaut assistant. Robots and intelligent machines in homes, at work sites, and in automated warehouses could also benefit from an autonomous object and action association ability. The only constraints on the research question of this thesis were the presence of task sequences and shared human-robot workspace. For such work environments, it can be concluded that the affordance-based task communication presented is a feasible and effective alternative method for requesting tasks, in addition to the explicit task requests.

7 Future work

The possibility of communicating a task by using only action or target object reference makes several novel applications available. For example, a “pointing only” interface could be used with affordances to give tasks to the robot, because merely pointing at an object could be translated into a task. If this approach is combined with some “yes or no” type of confirmation mechanism, it would be a viable way to communicate with the robot. The advantage of this approach is that humans are not required to remember target object names. The disadvantage, however, is that the robot can only operate with objects that it already knows or can recognise automatically.

Another potential application is automatic mission execution monitoring. Because the robot can predict the possible tasks using a partial speech input, it is possible to request confirmation of task requests that seem very unlikely, and to propose alternative tasks for execution. As the prediction is essentially based on knowledge of past sequences, the robot can adapt to any types of changes in the task sequences by just performing those sequences.

Some task requests could also include other parameters in addition to the action and target object references. For example, a “rotate antenna” task request could also include the angle to be rotated as a parameter. Requesting tasks with this kind of additional parameters could thus be one interesting direction to be researched. It is not self-evident that affordance-based task communication would still be beneficial in this case, because the experiments presented in this thesis did not include task-related parameters other than object and action names.

An automatic configuration of dialogue managers, when new devices are inserted into a network of devices, is one potential area of research for affordance-based task

communication. This is due to the fact that an affordance-based task communication configuration requires only information about actions that the devices can perform in order to be functional. There have already been plug-and-play interface systems where devices can automatically transfer dialogue information to the dialogue manager [109]. Affordance-based task communication could be examined as part of such a system.

References

- [1] J. A. Adams (2002). Critical considerations for human-robot interface development. In *Proc. of the AAAI Fall Symposium on Human-Robot Interaction*. Cape Cod, MA, USA.
- [2] H. Aghajan, J. Augusto, and R. Delgado (2009). *Human-Centric Interfaces for Ambient Intelligence*. Academic Press.
- [3] D. A. Allport (1987). *Perspectives on Perception and Action*, chapter Selection for action: some behavioral and neurophysiological considerations of attention and action, pp. 395–419. Lawrence Erlbaum Associates.
- [4] R. O. Ambrose, H. Aldridge, R. S. Askew, R. R. Burrige, W. Bluethmann, M. Diftler, C. Lovchik, D. Magruder, and F. Rehnmark (2000). Robonaut: NASA’s space humanoid. *IEEE Intelligent Systems and their Applications*, vol. 15:pp. 57–63.
- [5] A. Atrash, R. Kaplow, J. Villemure, R. West, H. Yamani, and J. Pineau (2009). Development and validation of a robust speech interface for improved human-robot interaction. *International Journal of Social Robotics*, vol. 1(4):pp. 345–356.
- [6] M. Bauer, G. Kortuem, and Z. Segall (1999). Where are you pointing at? A study of remote collaboration in a wearable videoconference system. In *Proc. of the 3rd International Symposium on Wearable Computers (ISWC)*, pp. 151–158. San Francisco, CA, USA.
- [7] D. Billman, M. Feary, and J. Zumbado (2011). Evidence report: risk of inadequate design of human and automation/robotic integration. *Technical report*, Lyndon B. Johnson Space Center, National Aeronautics and Space Administration (NASA).

- [8] R. Bolt (1980). Put-that-there: voice and gesture at the graphics interface. In *Proc. of the 7th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 262–270. Seattle, WA, USA.
- [9] G. Brat, M. Gheorghiu, D. Giannakopoulou, and C. Pasareanu (2008). Verification of plans and procedures. In *Proc. of the IEEE Aerospace Conference*. Big Sky, MT, USA.
- [10] T. Brick and M. Scheutz (2007). Incremental natural language processing for HRI. In *Proc. of the ACM/IEEE International Conference on Human-Robot Interaction*, pp. 263–270. Washington D.C., USA.
- [11] A. Brooks and C. Breazeal (2006). Working with robots and objects: revisiting deictic reference for achieving spatial common ground. In *Proc. of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, pp. 297–304. Salt Lake City, UT, USA.
- [12] R. Burridge and J. Graham (2002). Providing robotic assistance during extravehicular activity. In *Proc. of the SPIE: Mobile Robots XVI*, vol. 4573, pp. 22–33. Boston, MA, USA.
- [13] R. Burridge, J. Graham, K. Shillcutt, R. Hirsh, and D. Kortenkamp (2003). Experiments with an EVA assistant robot. In *Proc. of the 7th International Symposium on Artificial Intelligence, Robotics and Automation in Space (iSAIRAS)*. Nara, Japan.
- [14] N. Cabrol, J. Kosmo, R. Trevino, and H. Thomas (1999). Results of the 1st astronaut-rover (ASRO) interaction field experiment and recommendations for future planetary surface exploration. In *Proc. of the 18th Digital Avionics Systems Conference (DASC)*, vol. 2. Saint Louis, MO, USA.
- [15] S. Chong, Y. Kuno, N. Shimada, and Y. Shirai (2000). Human-robot interface based on speech understanding assisted by vision. In T. Tan, Y. Shi,

- and W. Gao (eds.), *Advances in Multimodal Interfaces (ICMI)*, vol. 1948 of *Lecture Notes in Computer Science*, pp. 16–23. Springer Berlin / Heidelberg.
- [16] W. J. Clancey (2002). Simulating activities: relating motives, deliberation, and attentive coordination. *Cognitive Systems Research*, vol. 3:pp. 471–499.
- [17] W. J. Clancey (2004). Roles for agent assistants in field science: understanding personal projects and collaboration. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 34:pp. 125–137.
- [18] W. J. Clancey, M. Sierhuis, R. Alena, D. Berrios, J. Dowding, J. S. Graham, K. S. Tyree, R. L. Hirsh, W. B. Garry, and A. Semple (2005). Automating capcom using mobile agents and robotic assistants. In *Proc. of the 1st Space Exploration Conference*. Orlando, FL, USA.
- [19] W. J. Clancey, M. Sierhuis, R. Alena, J. Dowding, M. Scott, and R. van Hoof (2006). Power agents: the mobile agents 2006 field test at MDRS. In *Proc. of the 9th International Mars Society Convention*. Washington D.C., USA.
- [20] H. H. Clark and S. E. Brennan (1991). *Perspectives on Socially Shared Cognition*, chapter Grounding in communication, pp. 127–149. American Psychological Association (APA) Books.
- [21] C. Culbert, J. Rochlis, F. Rehnmark, D. Kortenkamp, K. Watson, R. Ambrose, R. Diftler, B. Ward, L. Pedersen, and C. Weisbin (2003). Activities of the NASA exploration team human-robotics working group. In *Proc. of the Space 2003 Conference*. Long Beach, CA, USA.
- [22] K. Dautenhahn, S. Woods, C. Kaouri, M. L. Walters, K. L. Koay, and I. Werry (2005). What is a robot companion - friend, assistant or butler? In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1192–1197. Edmonton, Alberta, Canada.

- [23] B. D. Davison and H. Hirsh (1998). Predicting sequences of user actions. In *Proc. of the AAAI/ICML Workshop on Predicting the Future: AI Approaches to Time-Series Problems*, pp. 5–12. Madison, WI, USA.
- [24] M. Diftler, R. Ambrose, W. Bluethmann, F. Delgado, E. Herrera, J. Kosmo, B. Janoiko, B. Wilcox, J. Townsend, J. Matthews, T. W. Fong, M. Bualat, S. Y. Lee, J. Dorsey, and W. Doggett (2007). Crew/robot coordinated planetary EVA operations at a lunar base analog site. In *Proc. of the 38th Lunar and Planetary Science Conference (LPSC)*. League City, TX, USA.
- [25] M. Diftler, R. Ambrose, S. Goza, K. Tyree, and E. Huber (2005). Robonaut mobile autonomy: initial experiments. In *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1425–1430. Barcelona, Spain.
- [26] M. Diftler, C. Culbert, R. Ambrose, R. Platt, and W. Bluethmann (2003). Evolution of the NASA/DARPA Robonaut control system. In *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, vol. 2, pp. 2543–2548. Taipei, Taiwan.
- [27] P. Dillenbourg, M. Baker, A. Blaye, and C. O’Malley (1996). *The Evolution of Research on Collaborative Learning*, chapter Learning in humans and machine: towards an interdisciplinary learning science, pp. 189–211. Oxford: Elsevier.
- [28] P. Drews and P. Fromm (1997). A natural language processing approach for mobile service robot control. In *Proc. of the 23rd International Conference on Industrial Electronics, Control and Instrumentation (IECON)*, vol. 3, pp. 1275–1277. New Orleans, LA, USA.
- [29] J. Duchan (1995). *Deixis in Narrative: A Cognitive Science Perspective*, chapter Preschool children’s introduction of characters into their oral stories: evidence for deictic organization of first narratives, pp. 227–241. Lawrence Erlbaum Associates.

- [30] M. Duke, S. Hoffman, and K. Snook (2003). The lunar surface reference mission: a description of human and robotic surface activities. *Technical Report NASA TP-2003-210793*, NASA Lyndon B. Johnson Space Center (JSC), Houston, TX, USA.
- [31] J. Ferketic, L. Goldblatt, E. Hodgson, S. Murray, R. Wichowski, A. Bradley, T. W. Fong, J. Evans, W. Chun, R. Stiles, M. Goodrich, and A. Steinfeld (2006). Toward human-robot interface standards I: use of standardization and intelligent subsystems for advancing human-robotic competency in space exploration. In *Proc. of the SAE 36th International Conference on Environmental Systems (ICES)*. Norfolk, VA, USA.
- [32] J. Ferketic, L. Goldblatt, E. Hodgson, S. Murray, R. Wichowski, A. Bradley, T. W. Fong, J. Evans, W. Chun, R. Stiles, M. A. Goodrich, A. Steinfeld, D. King, and C. Erkorkmaz (2006). Toward human-robot interface standards II: an examination of common elements in human-robot interaction across the space enterprise. In *Proc. of the AIAA Space Conference*. San Jose, CA, USA.
- [33] G. Ferretti, G. Magnani, P. Putz, and P. Rocco (1996). The structured design of an industrial robot controller. *Control Engineering Practice*, vol. 4(2):pp. 239–249.
- [34] D. Ferrucci, E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. A. Kalyanpur, A. Lally, J. W. Murdock, E. Nyberg, J. Prager, et al. (2010). Building Watson: an overview of the DeepQA project. *AI Magazine*, vol. 31(3):pp. 59–79.
- [35] T. W. Fong (2001). *Collaborative Control: A Robot-Centric Model for Vehicle Teleoperation*. Ph.D. thesis, Robotics Institute, Carnegie Mellon University (CMU), Pittsburgh, PA, USA.
- [36] T. W. Fong and I. Nourbakhsh (2005). Interaction challenges in human-robot space exploration. *ACM Interactions*, vol. 12(2):pp. 42–45.

- [37] T. W. Fong, I. Nourbakhsh, and K. Dautenhahn (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, vol. 42(3-4):pp. 143–166.
- [38] T. W. Fong, I. Nourbakhsh, C. Kunz, L. Fluckiger, and J. Schreiner (2005). The peer-to-peer human-robot interaction project. In *Proc. of the AIAA Space 2005 Conference*. Long Beach, CA, USA.
- [39] J. P. Fritz, T. P. Way, and K. E. Barner (1996). Haptic representation of scientific data for visually impaired or blind persons. In *Proc. of the 11th Annual CSUN Conference on Technology and Persons with Disabilities*. Los Angeles, CA, USA.
- [40] P. Fromm and P. Drews (1998). Natural language processing for dynamic environments. In *Proc. of the 24th Annual Conference of the IEEE Industrial Electronics Society (IECON)*, vol. 4, pp. 2018–2021. Aachen, Germany.
- [41] K. Funakoshi, M. Nakano, T. Torii, Y. Hasegawa, H. Tsujino, N. Kimura, and N. Iwahashi (2007). Robust acquisition and recognition of spoken location names by domestic robots. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1435–1440. San Diego, CA, USA.
- [42] G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais (1987). The vocabulary problem in human-system communication. *Communications of the ACM*, vol. 30(11):pp. 964–971.
- [43] W. B. Garry, W. J. Clancey, M. X. Sierhuis, J. S. Graham, R. L. Alena, J. Dowding, and A. Semple (2005). Human-robotic field relations for the Moon: lessons from simulated Martian EVAs. In *Proc. of the Space Resources Roundtable VII: LEAG Conference on Lunar Exploration*. League City, TX, USA.

- [44] V. Gawron (2000). *Human Performance Measures Handbook*. Lawrence Erlbaum Associates.
- [45] J. J. Gibson (1977). *Perceiving, Acting and Knowing*, chapter The theory of affordances, pp. 67–82. Lawrence Erlbaum Associates.
- [46] M. A. Goodrich and A. C. Schultz (2007). Human-robot interaction: a survey. *Foundations and Trends in Human-Computer Interaction*, vol. 1:pp. 203–275.
- [47] J. Grezes and J. Decety (2002). Does visual perception of object afford action? Evidence from a neuroimaging study. *Neuropsychologia*, vol. 40(2):pp. 212–222.
- [48] M. Gullberg (1999). Gestures in spatial descriptions. *Lund Working Papers in Linguistics*, vol. 47:pp. 87–97.
- [49] A. Halme, I. Leppänen, J. Suomela, S. Ylönen, and I. Kettunen (2003). WorkPartner: interactive human-like service robot for outdoor applications. *The International Journal of Robotics Research*, vol. 22(7-8):pp. 627–640.
- [50] A. Hampapur, A. Senio, S. Pankanti, Y. Tian, G. Pingali, and R. Bolle (2002). Autonomic user interface. *Research Report RC22542*, IBM.
- [51] M. Hanheide, M. Hanheide, N. Hofemann, and G. Sagerer (2006). Action recognition in a wearable assistance system. In *Proc. of the 18th International Conference on Pattern Recognition (ICPR)*, vol. 2, pp. 1254–1258. Hong Kong.
- [52] S. Hart and L. Staveland (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Human Mental Workload*, vol. 1:pp. 139–183.
- [53] M. Hartmann (2010). *Context-Aware Intelligent User Interfaces for Supporting System Use*. Ph.D. thesis, Technische Universität Darmstadt, Germany.

- [54] S. S. Heikkilä (2010). Implementing human-robot interaction applications with GIMnet/MaCI. In *Proc. of the GIMNET 2010*. Espoo, Finland.
- [55] S. S. Heikkilä (2010). The role of natural interaction in astronaut-robot cooperation. In *Proc. of the International Astronautical Congress (IAC)*. Prague, Czech Republic.
- [56] S. S. Heikkilä, F. Didot, and A. Halme (2008). Centaur-type service robot technology assessment for astronaut assistant development. In *Proc. of the 10th ESA Workshop on Advanced Space Technologies for Robotics and Automation (ASTRA)*. Noordwijk, The Netherlands.
- [57] S. S. Heikkilä and A. Halme (2011). Indirect human-robot task communication using affordances. In *Proc. of the 20th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. Atlanta, GA, USA.
- [58] S. S. Heikkilä, A. Halme, and A. Schiele (2010). Human-human inspired task and object definition for astronaut-robot cooperation. In *Proc. of the 10th International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS)*. Sapporo, Japan.
- [59] P. Heiskanen, S. S. Heikkilä, and A. Halme (2008). Development of a dynamic mobile robot simulator for astronaut assistance. In *Proc. of the 10th ESA Workshop on Advanced Space Technologies for Robotics and Automation (ASTRA)*. Noordwijk, The Netherlands.
- [60] Y. Hirata, Z. Wang, and K. Kosuge (2006). Human-robot interaction based on passive robotics. In *Proc. of the 1st SICE-ICASE International Joint Conference*, pp. 4206–4209. Busan, South Korea.
- [61] S. J. Hoffman (2001). The Mars surface reference mission: a description of human and robotic surface activities. *Technical Report NASA TP-2001-209371*, NASA Lyndon B. Johnson Space Center (JSC), Houston, USA.

- [62] K. Holtzblatt and S. Jones (1993). *Participatory Design: Principles and Practices*, chapter Contextual inquiry: a participatory technique for system design, pp. 177–210. Lawrence Erlbaum Associates.
- [63] C. Hu, M. Meng, P. Liu, and X. Wang (2003). Visual gesture recognition for human-machine interface of robot teleoperation. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, vol. 2, pp. 1560–1565. Las Vegas, NV, USA.
- [64] X. Huang, F. Alleva, H. Hon, M. Hwang, and R. Rosenfeld (1993). The SPHINX-II speech recognition system: an overview. *Computer, Speech and Language*, vol. 7:pp. 137–148.
- [65] T. Iio, M. Shiomi, K. Shinozawa, T. Miyashita, T. Akimoto, and N. Hagita (2009). Lexical entrainment in human-robot interaction: can robots entrain human vocabulary? In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3727 –3734. Saint Louis, MO, USA.
- [66] H. Jones and S. Rock (2002). Dialogue-based human-robot interaction for space construction teams. In *Proc. of the IEEE Aerospace Conference*, vol. 7, pp. 3645–3653. Big Sky, MT, USA.
- [67] N. Kanas and D. Manzey (2008). *Space Psychology and Psychiatry*. Springer Verlag, 2nd edition.
- [68] L. Karttunen and S. Peters (1979). Conventional implicature. *Syntax and Semantics*, vol. 11:pp. 1–56.
- [69] T. Kaupp (2008). *Probabilistic Human-Robot Information Fusion*. Ph.D. thesis, University of Sydney, Australia.
- [70] I. Kauppi (2003). *Intermediate Language for Mobile Robots. A Link Between the High-level Planner and Low-level Services in Robots*. Ph.D. thesis, Helsinki University of Technology (TKK), Espoo, Finland.

- [71] C. C. Kemp, C. D. Anderson, H. Nguyen, A. J. Trevor, and Z. Xu (2008). A point-and-click interface for the real world: laser designation of objects for mobile manipulation. In *Proc. of the 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 241–248. Amsterdam, The Netherlands.
- [72] S. Kiesler (2005). Fostering common ground in human-robot interaction. In *Proc. of the IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN)*, pp. 729–734. Nashville, TN, USA.
- [73] V. Klingspor, J. Demiris, and M. Kaiser (1997). Human-robot communication and machine learning. *Applied Artificial Intelligence*, vol. 11(7):pp. 719–746.
- [74] G. Kminek (2004). Human Mars mission project: human surface operations on Mars. *Technical Report ESA Aurora/GK/EE/004.04*, European Space Research and Technology Centre (ESTEC) / European Space Agency (ESA), Noordwijk, The Netherlands.
- [75] C. Knipping (2008). A method for revealing structures of argumentations in classroom proving processes. *The International Journal on Mathematics Education (ZDM)*, vol. 40(3):pp. 427–441.
- [76] G. Kruijff, P. Lison, T. Benjamin, H. Jacobsson, and N. Hawes (2007). Incremental, multi-level processing for comprehending situated dialogue in human-robot interaction. In *Proc. of the Symposium on Language and Robotics*, pp. 55–64. Aveiro, Portugal.
- [77] V. Kulyukin (2004). Human-robot interaction through gesture-free spoken dialogue. *Autonomous Robots*, vol. 16(3):pp. 239–257.
- [78] W. Levelt, G. Richardson, and W. La Heij (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language*, vol. 24(2):pp. 133–164.

- [79] S. Li (2007). *Multi-Modal Interaction Management for a Robot Companion*. Ph.D. thesis, Bielefeld University, Germany.
- [80] F. Linton, D. Joy, H. Schaefer, and A. Charron (2000). OWL: a recommender system for organization-wide learning. *Educational Technology and Society*, vol. 3(1):pp. 62–76.
- [81] T. W. Malone and K. Crowston (1994). The interdisciplinary study of coordination. *ACM Computing Surveys*, vol. 26(1):pp. 87–119.
- [82] D. McNeill (1996). *Hand and Mind: What Gestures Reveal About Thought*. University of Chicago Press.
- [83] J. McPhee and J. Charles (2009). *Human Health and Performance Risks of Space Exploration Missions: Evidence Reviewed by the NASA Human Research Program*. U.S. Government Printing Office.
- [84] M. F. McTear (2002). Spoken dialogue technology: enabling the conversational user interface. *ACM Computing Surveys (CSUR)*, vol. 34(1):pp. 90–169.
- [85] J. Mehling, P. Strawser, L. Bridgwater, W. Verdeyen, and R. Rovekamp (2007). Centaur: NASA’s mobile humanoid designed for field work. In *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2928–2933. Rome, Italy.
- [86] Merriam-Webster Online Dictionary (2011). Interaction — Merriam-Webster Online. <http://www.m-w.com/dictionary/interaction>. [Online; accessed 19-October-2011].
- [87] L. Mignonneau and C. Sommerer (2005). Designing emotional, metaphoric, natural and intuitive interfaces for interactive art, edutainment and mobile communications. *Computers & Graphics*, vol. 29(6):pp. 837–851.

- [88] P. Milgram and F. Kishino (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information Systems*, vol. 77:pp. 1321–1329.
- [89] D. Moore, I. Essa, and M. Hayes (1999). Exploiting human actions and object context for recognition tasks. In *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, vol. 1, pp. 80–86. Bombay, India.
- [90] R. Moratz and T. Tenbrink (2008). Affordance-based human-robot interaction. *Lecture Notes in Artificial Intelligence (LNAI)*, vol. 4760:pp. 63–76.
- [91] P. Mulgaonkar, H. Dobbs, J. Blair, R. Dodd, M. Hofmann, D. Martinez, C. Mitchell, and R. J. Perna (2002). Ad hoc study on human robot interface issues. *Technical report*, Army Science Board (ASB), Department of Defense, United States Army, Arlington, VA, USA.
- [92] A. M. Naghsh, J. Gancet, A. Tanoto, J. Penders, C. R. Roast, and M. Ilzkovitz (2008). Human robot interaction in guardians. In *Proc. of the EURON/IARP International Workshop on Robotics for Risky Interventions and Surveillance of the Environment*. Benicassim, Spain.
- [93] NASA (2006). Lunar Exploration Objectives. National Aeronautics and Space Administration (NASA) document, http://www.nasa.gov/mision_pages/exploration/mmb/why_moon_process.html. [Online; accessed 18-October-2011].
- [94] A. Naumann, J. Hurtienne, J. Israel, C. Mohs, M. Kindsmüller, H. Meyer, and S. Hußlein (2007). Intuitive use of user interfaces: defining a vague concept. *Engineering Psychology and Cognitive Ergonomics*, pp. 128–136.
- [95] M. Neerinx, A. Bos, A. Olmedo-Soler, U. Brauer, L. Breebaart, N. Smets, J. Lindenberg, T. Grant, and M. Wolff (2008). The mission execution crew assistant: improving human-machine team resilience for long duration missions. In *Proc. of the 59th International Astronautical Congress (IAC)*.

- [96] D. Norman (1988). *The Design of Everyday Things*. Doubleday Business, New York, USA.
- [97] S. O’Keefe (2004). The vision for space exploration. *Technical Report NP-2004-01-334-HQ*, National Aeronautics and Space Administration (NASA), Washington D.C., USA.
- [98] C. Okoli and K. Schabram (2010). A guide to conducting a systematic literature review of information systems research. *Sprouts: Working Papers on Information Systems*, vol. 10(26):pp. 1–49.
- [99] M. S. Pandit and S. Kalbag (1998). The selection recognition agent: instant access to relevant information and operations. *Knowledge-Based Systems*, vol. 10(5):pp. 305–310.
- [100] J. Parker (2008). Buttons, simplicity, and natural interfaces. *Loading...*, vol. 2(2).
- [101] L. Pedersen, D. Kortenkamp, D. Wettergreen, I. Nourbakhsh, and T. Smith (2002). NASA EXploration Team (NEXT) space robotics technology assessment report. *Technical Report NASA*, Computational Sciences Division, NASA Ames Research Center, California, USA.
- [102] D. Perzanowski, A. Schultz, and W. Adams (1998). Integrating natural language and gesture in a robotics domain. In *Proc. of the IEEE International Symposium on Intelligent Control (ISIC)*, pp. 247–252. Gaithersburg, MD, USA.
- [103] D. Perzanowski, A. Schultz, W. Adams, and E. Marsh (1999). Goal tracking in a natural language interface: towards achieving adjustable autonomy. In *Proc. of the IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA)*, pp. 208–213. Monterey, CA, USA.

- [104] P. Peursum (2005). *Using Human Activity to Indirectly Recognise Objects in Indoor Wide-Angle Scenes*. Ph.D. thesis, Curtin University of Technology, Australia.
- [105] J. Pires (2005). Robot-by-voice: experiments on commanding an industrial robot using the human voice. *Industrial Robot: An International Journal*, vol. 32(6):pp. 505–511.
- [106] Princeton University (2011). Assistant — WordNet: an Electronic Lexical Database. <http://wordnetweb.princeton.edu/perl/webwn?s=assistant>. [Online; accessed 18-October-2011].
- [107] P. Putz and A. Elfving (1992). Control techniques 2, automation and robotics control development methodology definition report. *Technical Report ESA CT2/CDR/DO*, Dornier and European Space Research and Technology Centre (ESTEC) / ESA, Noordwijk, The Netherlands.
- [108] R Development Core Team (2011). *R: a Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [109] M. Rayner, I. Lewin, G. Gorrell, and J. Boye (2001). Plug and play speech understanding. In *Proc. of the 2nd SIGdial Workshop on Discourse and Dialogue*. Aalborg, Denmark.
- [110] A. Richter and P. Putz (2002). Automation and robotics for human Mars exploration (AROMA) final report. *Technical report*, Kayser-Threde GmbH and European Space Agency (ESA), Munich, Germany.
- [111] M. T. Rosenstein, A. H. Fagg, S. Ou, and R. A. Grupen (2005). User intentions funneled through a human-robot interface. In *Proc. of the 10th International Conference on Intelligent User Interfaces*, pp. 257–259. San Diego, CA, USA.

- [112] E. Rukzio (2006). *Physical Mobile Interactions: Mobile Devices as Pervasive Mediators for Interactions with the Real World*. Ph.D. thesis, Ludwig-Maximilians-Universität München (LMU), Munich, Germany.
- [113] J. Saarinen, A. Maula, R. Nissinen, H. Kukkonen, J. Suomela, and A. Halme (2007). GIMnet - infrastructure for distributed control of generic intelligent machines. In *Proc. of the 13th IASTED International Conference on Robotics and Applications Telematics*. Würzburg, Germany.
- [114] C. Sagan and R. Reddy (1979). Machine intelligence and robotics: report of the NASA study group - executive summary. *Technical report*, Jet Propulsion Laboratory (JPL), National Aeronautics and Space Administration (NASA), Pasadena, CA, USA.
- [115] G. Salvendy (1987). *Handbook of Human Factors*. John Wiley & Sons, New York, USA.
- [116] P. Salvini, C. Laschi, and P. Dario (2006). From robotic tele-operation to tele-presence through natural interfaces. In *Proc. of the 1st IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechanics (BioRob)*, pp. 408–413. Pisa, Italy.
- [117] O. C. Schrempf, D. Albrecht, and U. D. Hanebeck (2007). Tractable probabilistic models for intention recognition based on expert knowledge. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1429–1434. San Diego, CA, USA.
- [118] B. Sellner, F. Heger, L. Hiatt, R. Simmons, and S. Singh (2006). Coordinated multiagent teams and sliding autonomy for large-scale assembly. *Proceedings of the IEEE*, vol. 94(7):pp. 1425–1444.
- [119] R. Simmons, S. Singh, F. Heger, L. M. Hiatt, S. C. Koterba, N. Melchior, and B. P. Sellner (2007). Human-robot teams for large-scale assembly. In

- Proc. of the NASA Science Technology Conference (NSTC)*. Adelphi, MD, USA.
- [120] M. Skubic, D. Perzanowski, S. Blisard, A. Schultz, W. Adams, M. Bugajska, and D. Brock (2004). Spatial language for human-robot dialogs. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 34(2):pp. 154–167.
- [121] N. Smets, M. Abbing, M. Neerincx, J. Lindenberg, and H. van Oostendorp (2008). Game-based evaluation of personalized support for astronauts in long duration missions. In *Proc. of the 59th International Astronautical Congress (IAC)*. Glasgow, Scotland.
- [122] D. Stanley (2005). NASA’s exploration systems architecture study (ESAS), final report. *Technical Report NASA-TM-2005-214062*, The National Aeronautics and Space Administration (NASA).
- [123] A. Stoica, D. Keymeulen, A. Csaszar, Q. Gan, T. Hidalgo, J. Moore, J. Newton, S. Sandoval, and J. Xu (2005). Humanoids for lunar and planetary surface operations. In *Proc. of the IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, pp. 2649–2654. Hawaii, HI, USA.
- [124] K. Stubbs, D. Wettergreen, and P. Hinds (2007). Autonomy and common ground in human-robot interaction: a field study. *IEEE Intelligent Systems*, vol. 22:pp. 42–50.
- [125] K. Stubbs, D. Wettergreen, and I. Nourbakhsh (2008). Using a robot proxy to create common ground in exploration tasks. In *Proc. of the 3rd ACM/IEEE International Conference on Human Robot Interaction (HRI)*, pp. 375–382. Amsterdam, The Netherlands.
- [126] S. Sugano and T. Ogata (1996). Emergence of mind in robots for human interface - research methodology and robot model. In *Proc. of the IEEE*

- International Conference on Robotics and Automation (ICRA)*, vol. 2, pp. 1191–1198. Minneapolis, MN, USA.
- [127] O. Sugiyama, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita (2007). Natural deictic communication with humanoid robots. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1441–1448. San Diego, CA, USA.
- [128] O. Sugiyama, T. Kanda, M. Imai, H. Ishiguro, N. Hagita, and Y. Anzai (2006). Human-like conversation with gestures and verbal cues based on a three-layer attention-drawing model. *Connection Science*, vol. 18(4):pp. 379–402.
- [129] J. Suomela (2004). *From Teleoperation to the Cognitive Human-Robot Interface*. Ph.D. thesis, Helsinki University of Technology (TKK), Espoo, Finland.
- [130] W. Takano, K. Yamane, T. Sugihara, K. Yamamoto, and Y. Nakamura (2006). Primitive communication based on motion recognition and generation with hierarchical mimesis model. In *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3602–3609. Minneapolis, MN, USA.
- [131] P. Tarapore, M. Neibert, P. Tarapore, K. Biholar, J. Colombo, G. Linnell, H. Pant, and C. Underkoffler (2011). ATIS Telecom Glossary 2011. Alliance for Telecommunications Industry Solutions (ATIS) Document, <http://www.atis.org/glossary>. [Online; accessed 18-October-2011].
- [132] T. Tenbrink (2003). Communicative aspects of human-robot interaction. In H. Metslang and M. Rannut (eds.), *Languages in Development*. Lincom Europa.
- [133] S. Thrun (2004). Toward a framework for human-robot interaction. *Journal of Human-Computer Interaction*, vol. 19:pp. 9 – 24.

- [134] J. Trafton, N. Cassimatis, M. Bugajska, D. Brock, F. Mintz, and A. Schultz (2005). Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, vol. 35(4):pp. 460–470.
- [135] M. Tucker and R. Ellis (2004). Action priming by briefly presented objects. *Acta Psychologica*, vol. 116(2):pp. 185–203.
- [136] K. Tylén, M. Wallentin, and A. Roepstorff (2009). Say it with flowers! An fMRI study of object mediated communication. *Brain and language*, vol. 108(3):pp. 159–166.
- [137] B. Ullmer and H. Ishii (2000). Emerging frameworks for tangible user interfaces. *IBM Systems Journal*, vol. 39:pp. 915–931.
- [138] D. Vlasic, R. Adelsberger, G. Vannucci, J. Barnwell, M. Gross, W. Matusik, and J. Popovic (2007). Practical motion capture in everyday surroundings. In *Proc. of the SIGGRAPH Conference*, vol. 26. San Diego, CA, USA.
- [139] L. Wang, W. Hu, and T. Tan (2003). Recent developments in human motion analysis. *Pattern Recognition*, vol. 36(3):pp. 585–601.
- [140] M. Zebenay and S. S. Heikkilä (2010). Manipulator control for physical astronaut-robot interaction. In *Proc. of the 10th International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS)*. Sapporo, Japan.
- [141] E. Zereik, A. Sorbara, A. Merlo, E. Simetti, G. Casalino, and F. Didot (2011). Space robotics supporting exploration missions: vision, force control and coordination strategy for crew assistants. *Intelligent Service Robotics*, vol. 4(1):pp. 39–60.
- [142] V. W. Zue and J. R. Glass (2002). Conversational interfaces: advances and challenges. *Proceedings of the IEEE*, vol. 88(8):pp. 1166–1180.

Appendices

A Usage Examples of Control Development Methodology

This appendix describes with examples how ESA Control Development Methodology (CDM) was applied in Chapter 3. Table A.1 shows first how tasks were extracted based on overall astronaut-robot LAN setup mission description, which was given in Chapter 3, and how the extracted tasks were further converted to capability requirements. Then Table A.2 shows, using the TRANSPORT task as an example, how tasks were further extracted into subtasks.

Table A.1: Tasks of astronaut-robot LAN setup scenario extracted based on its overall mission description.

Task	Task description	Requirement description
1	Astronaut defines LAN setup mission (tools, components, etc.)	Astronaut can define mission scenarios (select tasks, actors, etc.)
2	Robot moves autonomously to the storage area	Robot can navigate and pilot autonomously
3	Robot identifies required objects at the storage area	Robot can localise and recognise static objects and areas
4	Robot identifies on-board storage locations for carrying objects	Robot can localise and recognise static objects and areas
5	Robot grasps the required objects	Robot can grasp objects
6	Robot inserts the objects for transfer on-board the rover	Robot can insert objects to defined locations
7	Robot moves autonomously to the target area	Robot can navigate and pilot autonomously
8	Astronaut identifies the exact place for LAN setup	Astronaut can point areas to the robot
9	Robot moves autonomously to the setup location	Robot can navigate and pilot autonomously

Table A.1: continues from the previous page...

10	Robot selects a specialised tool to prepare the installation location	Robot can use special tools
11	Robot uses a specialised tool to prepare the installation location	Robot can handle special tools
12	Robot selects manipulator to install the base station	Robot can handle special tools
13	Robot installs the base station to the prepared location	Robot can use special tools
14	Robot moves autonomously to the storage area	Robot can navigate and pilot autonomously
15	Robot finds storage containers for on-board objects	Robot can localise and recognise static objects and areas
16	Robot grasps the on-board objects	Robot can grasp objects
17	Robot inserts the objects into the storage containers	Robot can insert objects to defined locations

Table A.2: The TRANSPORT task analysed and extracted to subtasks.

Task	TRANSPORT TO <destination>
Examples	TRANSPORT TO geological_exploration_area; TRANSPORT TO storage_area
Definition	Move to a new destination. Wordreference: "move something or somebody around; usually over long distances".
Initial conditions	Subject in initial location. Initial and target end location known. Navigation and path planning available.
Boundary conditions	Do local and global path planning. Avoid collisions with environment. Maximum completion time.
Termination conditions	Subject in a desired end location.
Environment attributes	Navigation and obstacle avoidance procedures.
Subject attributes	Subject geometrical model. Piloting procedures.

B Software Architectures of the User Experiments

This appendix describes software architectures of the user experiments described in Chapter 4 and Chapter 5. The first section of the appendix presents shortly what is the GIM/Machine Control Interface (MaCI) software library that was used to build the experiments. The second and third sections describe the software architectures of user experiments having unambiguous and ambiguous task communication, respectively.

B.1 GIM/MaCI software library

The HRI systems in the thesis were built using the GIM/MaCI software library developed in the Automation Technology Laboratory at Aalto University [113]. The idea of the GIM/MaCI software is to provide a hardware abstraction layer that effectively makes components of same type look similar to the user. For instance, a user of the rangefinder MaCI module has to know only that the rangefinder returns a certain number of ranging measurements in order to use it. Application software that utilises one rangefinder, such as human localisation, should thus be directly suitable for any other possible rangefinder.

Another feature of the GIM/MaCI library is that the software is inherently modular because all the functionalities are separated to their own modules. This makes the code reusable, as it is easy to take out certain modules to be utilised in a new robot. Another equally significant benefit is the distributed computation load. The GIM/MaCI modules can run on any computer as long as they are able to connect to each other through a TCP/IP network.

B.2 User experiments with unambiguous task communication

The software architecture used for the user experiments having unambiguous task communication is shown in Figure B.1. The software was distributed primarily to two computers. A first computer was dedicated to receiving, sending, and processing of user interactions. The second was the computer on-board the robot, which created the MaCI server interfaces to the robot devices, such as the SICK rangefinder, and was controlling the behaviours of the robot, such as obstacle avoidance.

All the arrows between modules that have not been named in Figure B.1 use GIM/MaCI communication and are connected through a GIM/MaCI access point [113], which is a centralised router of the GIM network that enables, for instance, bypassing of company firewalls. A GIM access point enables modules also to announce the services that they can provide and share data to multiple clients. For example, when the rangefinder data is sent to the GIM access point, it is distributed directly to both Simultaneous Localisation and Mapping (SLAM) and human localisation clients.

B.3 User experiments with ambiguous task communication

The software architecture used for the user experiments having ambiguous task communication is shown in Figure B.2. This software was distributed primarily to two computers. The first one ran the speech recognition software while the second processed the dialogue and outputted the robot's speech utterances.

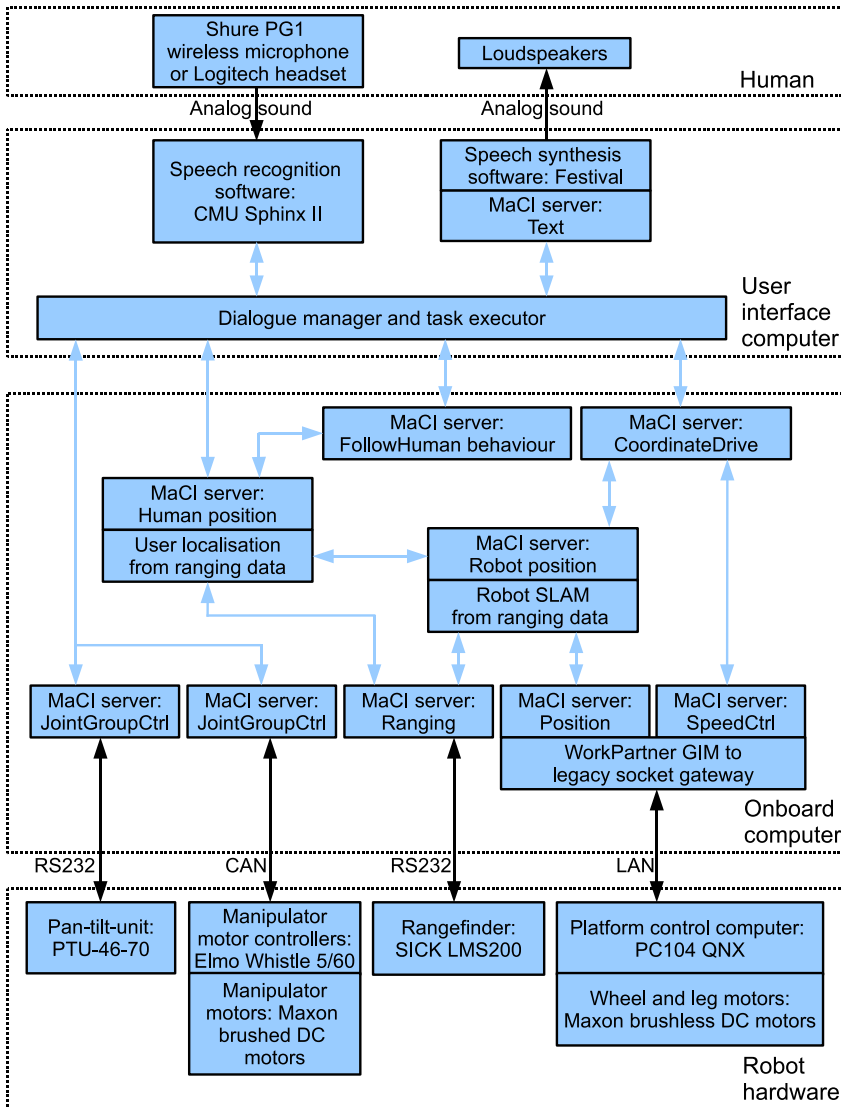


Figure B.1: Software architecture used in the user experiments that had unambiguous task communication. The arrows without labels indicate traffic through a GIM access point.

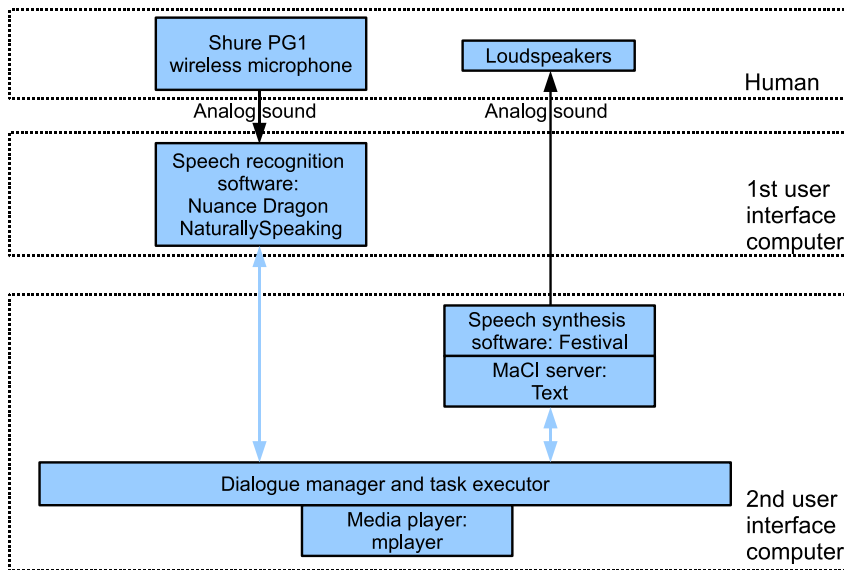


Figure B.2: Software architecture used in the user experiments that had ambiguous task communication. The arrows without labels indicate traffic through a GIM access point.

Efficient human-robot cooperation is currently hampered by a lack of truly human-compatible task communication methods. The suggested solution is indirect task communication based on the human-like ability to utilise affordances, i.e. action possibilities, in task communication. These so-called affordance-based task communication methods are compared with conventional direct task communication methods - in which all task parameters need to be communicated explicitly - and are shown to be able to simultaneously decrease human workload and task communication times. Furthermore, indirect and mixed direct/indirect task communication methods are found to be preferred over direct task communication methods. These findings show the feasibility and effectiveness of the approach in facilitating human-robot task communication that includes a priori known or recurring task sequences.

ISBN 978-952-60-4330-2 (pdf)

ISBN 978-952-60-4329-6

ISSN-L 1799-4934

ISSN 1799-4942 (pdf)

ISSN 1799-4934

Aalto University
School of Electrical Engineering
Department of Automation and Systems Technology
www.aalto.fi



**BUSINESS +
ECONOMY**

**ART +
DESIGN +
ARCHITECTURE**

**SCIENCE +
TECHNOLOGY**

CROSSOVER

**DOCTORAL
DISSERTATIONS**