Lehrstuhl für Steuerungs- und Regelungstechnik
Technische Universität München
Univ.-Prof. Dr.-Ing./Univ. Tokio Martin Buss

Alignment Strategies for Information Retrieval in Prosocial Human-Robot Interaction

Barbara Andrea Kühnlenz

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

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing. Klaus Diepold

Prüfer der Dissertation:

1. TUM Junior-Fellow Dr.-Ing. Dirk Wollherr
2. Univ.-Prof. Dr.-Ing. habil. Gerhard Rigoll
3. Univ.-Prof. Dr.-phil. habil. Hermann Müller, Ludwig-Maximilians-Universität München

Die Dissertation wurde am 19.06.2013 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 12.11.2013 angenommen.

Foreword

This dissertation has emerged from three years of work at the Institute of Automatic Control Engineering (LSR), Technische Universität München, where I stayed from 2010 until now. This interdisciplinary journey has been the greatest challenge in my life and broadened my mind forever.

My deepest thanks to my colleagues and friends, Dr. Andrea Bauer and Dr. Michelle Karg, for always encouraging me, for their open-mindedness, for sharing scientifical and non-scientifical ups and downs, and for the great time we spent together. Special thanks to my colleague Stefan Sosnowski, for his interdisciplinary interest and the pleasure to collaborate with him in our "EDDIE"- experiments. To Christian Landsiedel, for his amazing WOz-tool, Dr. Christoph Mayer (CS, TU München) and Jürgen Blume (MMK, TU München), for collaborative experiments. To Dr. Astrid Weiss and Nicole Mirnig (ICT&S Center, University of Salzburg), for their great collaboration, and for delivering me with delicious "Mozart Kugeln". To Prof. Björn Granström, Dr. Joakim Gustafson and Dr. Gabriel Skantze (TMH, KTH Stockholm), for the fruitful discussions and collaboration during my research stay at KTH, Stockholm. To Dr. Michael Zehetleitner, for inspiring discussions and collaboration. To my students, Katja Gütlein, Dennis Schenzow, Veronika Eder, in particular, Malte Buss, for sharing my research interests and constantly assisting me. To my room mates, Jens Hölldampf, for helping me patiently with his expertise in LaTeX and graphics-design at the beginning, Alper Ergin, for sharing instant coffee and chewing gums, and my "phantomous" always friendly room mate Vicente Garcia. To Dr. Angelika Peer, Antonia Glaser and Kati Landsiedel, for their advice and help in statistics. To Ken Friedl, for always taking excellent photos and video-tapes. And of course to the IURO-team, for sharing all ups and downs in the project, for taking over duties, making experiments work, and for making me always feel welcome in spite of my continuous absence from Lunch. To all my other LSR-team mates at our institute, for providing such a good working atmosphere. To the team of the TUM Institute for Advanced Study (IAS), for their indispensable financial and administrative support. I greatly appreciate the technical support from Mr. Jaschik, who patiently helped me with recovering my laptop from a crash, as well as of Mr. Stoeber, our admins, and the help in administrative issues from Mrs. Schmid, Mrs. Werner, and Mrs. Renner. To all my friends, thank you for your continuous empathy and patience.

In deepest gratitude I would like to thank my parents, for all their continuous support and encouragement. Most of all, I want to thank my husband Kolja for his love, care, and patience, for always being at my side, and for encouraging me whenever I needed it.

Munich, 2013. Barbara Kühnlenz

For my family

Abstract

The introduction of domestic robots into the real world faces a variety of interdisciplinary challenges. In particular, user acceptance and the willingness of humans to cooperate and interact with robots have to be maintained. In turn, robots should be aware of their situational knowledge limitations and be able to pro-actively and flexibly acquire the knowledge needed to perform their given objectives. This dissertation focuses on the development of informational and emotional alignment strategies to provide a prosocial common ground in human-robot interaction (HRI), and to pro-actively acquire missing task-knowledge from natural language dialog.

A robot has to cope with various environmental impacts, e.g. noisy outdoor conditions. In order to overcome this bottleneck of speech recognition, different dialog strategies, as well as specified miscommunication handling requests are developed and experimentally evaluated in this dissertation. In order to increase the efficiency of information retrieval in case of varying speech recognition performance while maintaining highest possible naturalness for the user, a switching mechanism is developed to allow smooth transitions between open and closed requests. The dialog strategies are embedded in a framework for pro-active information retrieval, implying hypothesis-driven information processing, as well as representing and evaluating the acquired knowledge during task-execution.

User acceptance and the willingness to cooperate and interact with a robot are increased by a targeted integration of social-psychological interaction mechanisms in a behavior control model. In human-human interaction, empathy and a feeling of having something in common with a person in need of help, e.g. in personal attitudes, are essential motivational influence factors for prosocial behavior. In this work, the developed mechanisms of emotional behavior control are successfully applied in evaluative experiments to induce these feelings in human users towards a robot. In particular, this is achieved by a combination of emotionally adaptive mimicry and speech, and pro-active small-talk mechanisms, employed prior to task-related interaction in order to establish a prosocial and cooperative common ground, generalizable to any human-robot interaction.

In order to combine the explored aspects of informational and emotional alignment into an integrated approach, a general framework is developed for task-related HRI, dividing a task in cognitive and social sub-tasks. The generalizability of the fully integrated approach is evaluated in an urban outdoor field trial, extended by an emotion recognition module to emotionally align with humans in a fully automated way during information retrieval. Since emotional alignment is highly associated with the legibility of emotional expressions, a differentiated assessment of the design-dependent legibility and user-acceptance issues is conducted in this thesis for emotional speech and mimicry, comparing two differently designed robotic heads of either machinelike versus more humanlike design. Additionally, impacts of dispositional empathy on the human performance in identifying the animated emotions are revealed, and the importance of an interactive context for emotion recognition is confirmed.

The experimental results of the outdoor field trial re-confirm the positive effects of combining informational and emotional alignment since all tested dimensions of user experience resulted in comparably high mean ratings, positively related with the willingness of humans to help a robot. Finally, the results indicate that informational and emotional

alignment are compensating each other with regard to user experience: In case of poor speech recognition performance in outdoor environments, successful emotional alignment keeps up the interest of humans to cooperate with a robot.

Summarizing, it can be deduced that proactive information retrieval benefits from an integration of emotional alignment strategies, since it reinforces the underlying prosocial motivation of humans to help a robot, thereby even compensating decreases in user acceptance due to bad speech recognition performance. Accordingly, the ideas, concepts, and approaches developed in this thesis significantly advance the state of the art in design and control of social HRI and information extraction from natural language dialogs.

Zusammenfassung

Der Einsatz von privaten Service-Robotern birgt eine Vielzahl von interdisziplinren Herausforderungen. Die Benutzerakzeptanz sowie die menschliche Bereitschaft, mit Robotern zu kooperieren und zu interagieren spielt dabei eine besondere Rolle. Im Gegenzug sollten Roboter die Grenzen ihres Wissens kennen und dazu im Stande sein, sich benöigtes Wissen auf pro-aktive und flexible Weise anzueignen, um die vom Menschen gestellten Aufgaben zu erfüllen. Der Schwerpunkt dieser Dissertation liegt in der Entwicklung von informationsbezogenen und emotionalen Anpassungsstrategien, um einerseits eine prosoziale Basis in der Mensch-Roboter Interaktion zu schaffen und andererseits Robotern einen pro-aktiven Informationsgewinn aus natürlichsprachlichen Dialogen zu ermöglichen.

Roboter müssen robust gegen viele Umwelteinflüsse werden, z.B. gegen hohe Lärmpegel in Außenbereichen. Zur Überwindung dieses Hindernisses für die Spracherkennung des Roboters, werden in dieser Dissertation Dialogstrategien und spezifische Rückfragen zur Klärung von Mißverständnissen entwickelt und experimentell evaluiert. Um die Effizienz der Informationsbeschaffung im Fall einer instabilen Spracherkennungsleistung zu erhöhen, und gleichzeitig die größtmögliche Natürlichkeit für den menschlichen Gesprächspartner aufrecht zu erhalten, wird ein Schaltmechanismus entwickelt, der fließende Übergänge zwischen offenen und geschlossenen Fragen im Dialog ermöglicht. Die Dialogstrategien sind in ein Rahmenkonzept für pro-aktive Informationsbeschaffung eingebettet, das sowohl hypothesengesteuerte Informationsverarbeitung vorsieht, also auch die Repräsentation und Evaluierung des akquirierten Wissens während der Ausführung einer Aufgabe.

Benutzerakzeptanz und menschliche Kooperationsbereitschaft werden in dieser Arbeit durch den Transfer von sozialpsychologischen Interaktionsmechanismen in die Verhaltenssteuerung eines Robotersystems erhöht. In zwischenmenschlichen Interaktionen sind Empathie, zusammen mit dem Gefühl, etwas mit der Hilfe benötigenden Person gemeinsam zu haben, z.B. gemeinsame persönliche Eigenschaften, essentielle Motivationsfaktoren für prosoziales Verhalten. In den evaluativen Experimenten dieser Arbeit, konnten die entwickelten Mechanismen zur emotionalen Verhaltenssteuerung erfolgreich eingesetzt werden, um diese Gefühle gegenüber einem Roboter zu induzieren. Im Einzelnen wird dies durch eine Kombination aus emotional angepasster Gesichtsmimik und Sprachprosodie mit pro-aktivem Small-Talk vor der aufgabenbezogenen Interaktion erreicht. Dadurch wird eine prosoziale und kooperative gemeinsame Gesprächsbasis geschaffen, die generalisierbar in jeder Mensch-Roboter Interaktion anwendbar ist.

Um die erforschten Aspekte von informationsbezogener und emotionaler Anpassung zu einem integrierten Ansatz zusammenzuführen, wird ein generisches Rahmenkonzept für aufgabenbezogene Mensch-Roboter Interaktion entwickelt, in dem eine Aufgabe sowohl in kognitive als auch in soziale Teilaufgaben eingeteilt wird. Die Generalisierbarkeit dieses integrativen Ansatzes wird in einem Feldexperiment in urbaner Umgebung evaluiert. Um während der Informationsdialoge eine voll automatisierte emotionale Anpassung an die Menschen vornehmen zu können, wird der Ansatz hierbei durch ein Emotionserkennungsmodul erweitert. Da emotionale Anpassung in einem engen Zusammenhang mit der Lesbarkeit emotionaler Ausdrücke steht, wird eine differenzierte Bewertung der designabhängigen Lesbarkeit unter Berücksichtigung von Benutzerakzeptanz für die emotionale Gesichtsmimik und Sprachprosodie zweier Roboterköpfe im Verlgeich mit maschinenhaftem versus menschenähnlicherem Aussehen durchgeführt. Zudem wird der Einfluss von

dispositionaler Empathie auf die menschliche Erkennungsleistung der animierten Emotionen aufgezeigt, und die Wichtigkeit eines interaktiven Kontextes für die Emotionserkennung bestätigt.

Die experimentellen Ergebnisse des Feldexperiments erbringen eine Rückbestätigung für die positiven Effekte einer Kombination aus informationsbezogener und emotionaler Anpassung, da die Mittelwerte der menschlichen Nutzererlebnisse auf allen getesteten Dimensionen verlgeichsweise hoch sind und zudem positiv mit der menschlichen Kooperationsbereitschaft mit einem Roboter korrelieren. Abschließend weisen die Ergebnisse darauf hin, dass informationsbezogene und emotionale Anpassung sich gegenseitig in Bezug auf das Nutzererlebnis kompensieren: Im Fall schlechter Spracherkennung durch das System in Außenbereichen kann erfolgreiche emotionale Anpassung an den Nutzer dessen Interesse zur Kooperation mit dem Roboter aufrechterhalten.

Zusammenfassend lässt sich sagen, dass pro-aktive Informationsbeschaffung von der Integration emotionaler Anpassung profitiert, da diese die zugrundeliegende Motivation zum prosozialen Verhalten in Menschen verstärkt und in Zuge dessen sogar sinkende Benutzerakzeptanz aufgrund schlechter Spracherkennung des Systems kompensieren kann. Folglich erhöhen die in dieser Arbeit entwickelten Ideen, Konzepte und Ansätze den Stand der Forschung maßgeblich in Bezug auf Design und Verhaltenssteuerung in der Mensch-Roboter Interaktion, sowie bezüglich der Informationsextraktion aus natürlichsprachlichen Dialogen.

Contents

Notations

Abbreviations & Symbols & Conventions

HRI	Human-Robot Interaction
HHI	Human-Human Interaction
HCI	Human-Computer Interaction
UX	user experience
DOF	degrees of freedom
WOz	Wizard-of-Oz
DS	dialog strategy
MHR	Miscommunication Handling Request
TPH	Theory of Perceptual Hypotheses
SDS	Spoken Dialog System
FSM	Finite State Machine
IR	Information Retrieval
SSD	Social Sub-dialog
EDDIE	Emotion Display with Dynamic Intuitive Expressions
IURO	Interactive Urban Robot
FACS	Facial Action Coding System
PAD	Pleasure-Arousal-Dominance
SMM	Social Motivation Model
UTAUT	Unifed Theory of Acceptance and Use of Technology
ANOVA	Analysis of Variance
MANOVA	Multivariate Analysis of Variance
SD	Standard Deviation
u'	Utterance
ε	Critical confidence threshold
α	Loss or gain in confidence/ statistical significance level
δ	Critical threshold of miscommunication
n	Number of subjects
F	Effect-size of ANOVA
T	Effect-size of T-tests
p	Significance
Cronbach's α	Statistical test for internal reliability of a questionnaire-construct

List of Figures

List of Tables

1 Introduction

Robots are more and more entering our daily lives and gradually moving from strictly structured industrial settings into our private households. Prominent examples of current commercial domestic robots are a vacuum cleaner and robotic toys, see Figure 1.1.

Currently, commercial robots are limited to one specific functionality and, accordingly, their design and capabilities correspond to this functionality, e.g., there is no need for a vacuum cleaner to be socially interactive, and the only purpose of robotic toys is to entertain human users. Thus, current domestic robots are mainly regarded as tools, differing in their function, and social interaction is limited to entertainment purposes. Nevertheless, recent studies show that humans tend to anthropomorphize these tools, i.e. by attributing live-like qualities to them [23]. One example for this effect is the use of anthropomorphic language when talking about technical devices. Thereby, the extend of anthropomorphic language is bound to the function and the design of the devices, i.e., a robotic dog is specifically more anthropomorphized than a robotic vacuum cleaner or an iPad [48].

In the research field of robotics, there is an increasing trend towards developing robots, that are designed to assist human users in more complex cognitive tasks and, thus, cover more than one functionality, thereby encountering for a combination of social and cognitive skills that go beyond the capabilities of tools. This process is comparable to human phylo- and ontogeny, providing evidence for initially separated genetic roots of cognition and speech with an originally isolated function: According to [159], human thinking was initially involved in the use of tools, and communication was solely associated with its social function in terms of utterances that are "directly related to the action itself", and not on a meta-level, where humans communicate *about* any topic. Thus, in the ontogenetic development of a child, a "pre-linguistic stage" in cognition, and a "pre-intellectual stage" of communication, is established. These different genetic lines are developing independently of each other, until a certain point is reached in infancy, "whereupon thought becomes verbal and speech rational" [159], as long as the full complexity of social cognition is developed and communicable.

For robotic research, this means that a combination of cognitive and social skills is indispensable when aiming for robotic "companions" [99, 118], designed to support and complement human users in socio-cognitive tasks of their daily lives, starting with possible differentiation between a task and its different sub-tasks, and/or the consideration of user preferences while cleaning the floor, up to coaching or rehabilitation applications. In every case, the difference to current commercial robotic vacuum cleaners and robotic toys is that human users will communicate with their robot *about* its task(s) or any other topic, which is not longer confined to an "action itself" and, thus, turns into meta-communication. This means, that human-robot communication evolves beyond simple speech commands that trigger a robotic action towards deeper understanding dialog systems, capable of extracting and representing task-relevant information from human speech input in ontological databases.

Figure 1.1: Examples of current commercial domestic robots: AIBO[1], Roomba[2] and PLEO[3]

As a consequence, a *need* of interaction evolves between humans and robots when sharing a more complex task-context, in contrast to *optional* playful interactions with robotic toys. Since communication is not restricted to explicit verbal utterances but also entails implicit non-verbal components, e.g., mimicry, extensive studies are conducted in the research field of human-robot interaction (HRI) to use and maximize the above mentioned effects of anthropomorphism in HRI [47]. The resulting interaction-designs for robots range from physical appearance to their interactive behavior in order to render the communication with them most intuitive for potential human users [6], or to enhance playful interactions by educational and/or therapeutic benefits for humans [41, 95, 114].

In human-human interaction (HHI), Watzlawick is often cited for his foundational observations on the nature of human communication: "One cannot *not* communicate", and "Every communication has a content and a relationship aspect such that the latter classifies the former and is therefore a meta communication" [163]. Both paradigms underline the socio-cognitive character of communication, i.e., an informational content is always embedded in the social context of an interaction and, thus, the ambiguity of natural language can only be interpreted in consideration of the social and situational context. Secondly, the "relationship aspect" is not only motivating the interpretation of informational contents, but also expressed in implicit non-verbal communication, meaning that humans even express their attitudes by conscious or unconscious body-language, or simply by not saying anything where a response is expected by an interaction partner. This goes in line with the findings of Spitz, who defined communication as "Each noticeable, conscious or unconscious, directed or undirected change of behavior [...], through which a human persuades willingly or unwillingly perception, feelings, affects and thoughts of others." [141] Based on these definitions it can be deduced that any communication is behavioral persuasion at the same time, and that informational and emotional communication contents are in a permanent interrelationship to each other during an interaction. Hence, in order to bring HRI to a socio-cognitive level, both, the informational and the emotional dimension of communication have to be considered and aligned with human interaction partners.

Within this context, this thesis explores informational and emotional alignment strategies and their benefit for the retrieval of missing task-knowledge from humans in prosocial HRI. The main challenges faced by the design and control of informational and emotional alignment are summarized in the following.

[1] www.sony-europe.com/aibo
[2] www.irobot.com/global/de/store/Roomba
[3] www.pleoworld.com

1.1 Challenges

The challenges faced in this thesis focus on generalizable approaches to solve informational and emotional alignment issues in the pro-active retrieval of missing task-knowledge from humans. The presented work is part of the FP7 STREP Interactive Urban Robot (IURO) project [1], a follow-on project of the Autonomous City Explorer (ACE)[2], where a robot has the task to find its way to a given goal location, e.g. because of being sent on errands by its human user. As a challenging example for missing task-knowledge, the robot does not posses any prior route knowledge or GPS. Accordingly, it has to find its way only by asking passers-by for the way, as can be seen in Figure 1.2.

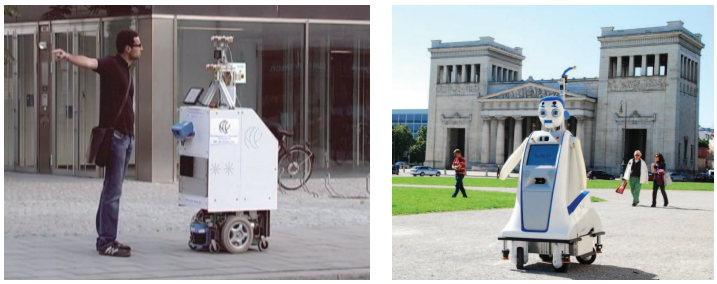

Figure 1.2: Exemplary outdoor interactions of the robots ACE and IURO

In the ACE project, the input modalities for humans were restricted to pointing gestures and buttons on a touchscreen. In the IURO project, the input modalities are enhanced by natural language speech input, as the most challenging modality of information retrieval in noisy outdoor environments. Another challenge is to motivate unconcerned passers-by to help the robot with missing route-knowledge for a task they do not benefit from.

Accordingly, this interdisciplinary approach faces several challenges for informational and emotional alignment in HRI. In contrast to indoor settings, informational alignment is impaired by poor speech recognition performance in outdoor environments, and thereby also affects user experience (UX). On the other hand, for emotional alignment the challenge is to set up a social situational context between a robot and humans in public spaces, where interactions can be easily interrupted by unforeseen environmental impacts. A main challenge is to trigger and keep up the motivation of human users to interact with the robot in a prosocial way and keep up their interest to provide it with the missing task-knowledge instead of walking away. Hence, a relation has to be created between the functionalities of cognitive and emotional interaction components in an autonomous system in order to embed information retrieval in a socially situated context. One way towards this goal is to simulate infant development in robots by employing learning and reinforcement techniques to build a social cognition, just like in child-adult interaction, e.g., [25]. However, learning is very time-consuming and although it should surely be integrated in domestic service robots, robots should be pre-equipped with a framework incorporating both, cognitive and social task-dependent guidelines to be able to align with their human users and, thus, be ready-to-use for their tasks in private households.

The key issues targeted in this thesis are summarized as follows:

[1] http://www.iuro-project.eu
[2] http://www.lsr.ei.tum.de/research/research-areas/robotics/ace-the-autonomous-city-explorer-project

Miscommunication

Since the communication quality in public spaces is often impaired by noisy environmental impacts, it is difficult for a robot to retrieve missing task-information from humans. The use of natural language for information retrieval entails a number of difficulties, e.g. vagueness and ambiguity of spoken language, and the technical challenges posed by automatic speech recognition. Thus, apart from background noise, miscommunication may be caused by lexical or conceptual difficulties. Most existing approaches, where robots ask humans for missing task-knowledge in order to extract information about the environment are still operating in very simple structured indoor environments [2, 89, 102, 106]. These robots are able to interpret and follow simple route instructions, but cannot cope with the complexity and vagueness of natural language. Several studies successfully explored miscommunication and informational complexity arising from users giving verbal route instructions in a simulated dialogic real-time interaction with an artificial agent executing the route instructions during the experiment [82, 135], and related error handling is integrated in spoken dialog systems, e.g. [137]. Moreover, in contrast to state-of-the-art approaches, this work addresses a robot that executes previously gained route instructions autonomously within real outdoor environments. Thus, complexity and the range of potential errors increases enormously, e.g. informational misalignment may be undetected during the dialog but leads to errors during execution of the gained route knowledge. Thus, it is necessary to represent and evaluate each newly gained information. As discontinuation due to weak speech recognition complicates information retrieval from humans in public spaces, proper handling of errors and miscommunication has to be ensured in order to provide validity of the acquired information on the one hand, and a satisfactory experience for the human interaction partners on the other hand.

User Acceptance

Another main challenge addressed in this thesis is the investigation of user acceptance - related issues. The perception of robots by human users is not always positive and may decrease over the time: First studies on the temporal progress of user experience (UX) in households equipped with a robotic vacuum cleaner indicate an initial enthusiasm in human users, but any enthusiasm may decrease over time due to habituation [14]. Interactive robots may even raise the initial enthusiasm [165], but some humans may be willing to explore the limits of robots, as observed in robotic applications developed to operate in public spaces, where even bullying behavior was shown by human passers-by towards the robot, e.g. [127]. In applications where useful information is provided by a robot, e.g. shopping recommendations in malls [73, 75], the interactions are initiated by human users, willing to interact with a robot for their own benefit. However, in this work the beneficial effect is reversed: A robot is proactively initiating an interaction with humans in order to get their help, i.e., to retrieve missing task-knowledge from them for its own benefit. Thus, also prosocial biases have to be explored in order to trigger and maintain a positive attitude in human users towards a robot. For example, Siegel et al. [134] could show that giving the robot a gender can be of high use to induce positive attitudes in human users. Numerous predictions on human behavior can be found in empirically validated theories from social psychology [51]. Hence, one main challenge is to investigate the transferability of these behavior predictions to HRI and, thus, make them integrable in robotic behavior control models.

Emotional Expressiveness

In order to involve emotional alignment in HRI, the emotional expressiveness of a robot has to be considered in terms of legibility-issues regarding intentions and social cues.

In human-human interaction (HHI), studies on unconscious mimicry present findings on the importance of facial mimicry in social interaction [34]. Thereby, the ability of interpreting emotional expressions plays a key role for feeling empathy towards others, as already developed in infants [19]. Dysfunctions might lead to social deficits, as observed in autism [40]. Also in HRI, emotion recognition, expression, and emotionally enriched communication and closed-loop behavior control have gained strong attention during the last two decades [80, 94, 112, 119, 129]. A number of studies have already been conducted which employ empathy as a factor in human-robot or human-computer interaction to manipulate the attitude of users towards an artificial agent, which can be categorized whether the artificial agents are used to express empathy [38, 99, 110, 113, 118, 149, 155] or induce it in the user [113, 114, 124]. Empathic expressions by the agents are mostly utilized to enhance the user experience and thus provide a benefit to the user. Another approach is to induce empathy in the user. This is, for example, achieved via facial mimicry [124]. The detection of emotions and its use in behavior control is treated in several works, e.g., e-learning systems [3], pedagogical agents [46], driver assistants [4], virtual agents [68], psychological assistance [72]. However, the effectiveness of automatic emotion recognition is still very limited and the connection between perceived and real emotions remains an open issue. In order to achieve the goal of incorporating social cues in interactive robots, there are extensive research efforts on building robot heads or robots with a full body that are able to express emotions, e.g. [29, 67, 151]. As related work shows, there are many different designs for emotionally expressive robots, whereas a differentiated validation of which is often pending. However, the legibility of the behavior of a robot is important for human users in order to interpret its intentions, and for social cues to take effect. This holds true for various behaviors, e.g., the legibility of the navigation behavior of a robot [92]. This thesis focuses the legibility of emotional robotic expressiveness in facial expressions and prosody in speech, as well as their impacts on task-relevant HRI.

Integration of Social and Cognitive Interaction Components

Prosocial HRI in the sense of proactively triggering effects of socialization between a robot and a human is not restricted to information transfer, accompanied by emotional expressions. In human-human interaction, socialization is not only established by implicit emotional expressions, but also in an explicit way. For example, human communication patterns show mechanisms of small talk, also called “phatic communication” to establish a common ground in form of shared beliefs between the interlocutors [17, 96]. Also many robots use small talk, but do not evaluate its specific influence on the interaction: Grace and George, two robots used as receptionist and guide at a conference [162], the seal robots used in elderly care [160] and Breazeal et al [24] use small talk for their robotic weight loss coach as a means for bonding and evaluation. Bartneck et al. [8] state phatic communication to be an important factor when judging the social abilities of a robot. This is consistent with Lee et al. [91], which evaluated human expectations when talking with a robot. Results show that people not only tend to greet human-like robots, but also use small talk with them during interaction instead of treating them like a non-social ticket automaton - even with no background knowledge about the abilities of the robot. Accordingly, a main challenge is not only to enrich dialog strategies for information retrieval by

emotional mimicry and prosody in speech, but also to model integrated architectures that combine cognitive tasks of information retrieval with social tasks to improve efficiency and naturalness of the interaction.

1.2 Main Contributions and Outline of the Thesis

The presented work focuses on informational and emotional alignment strategies to be combined in an integrative approach for prosocial information retrieval from natural language HRI, that give the structure to this thesis as highlighted in Figure 1.3.

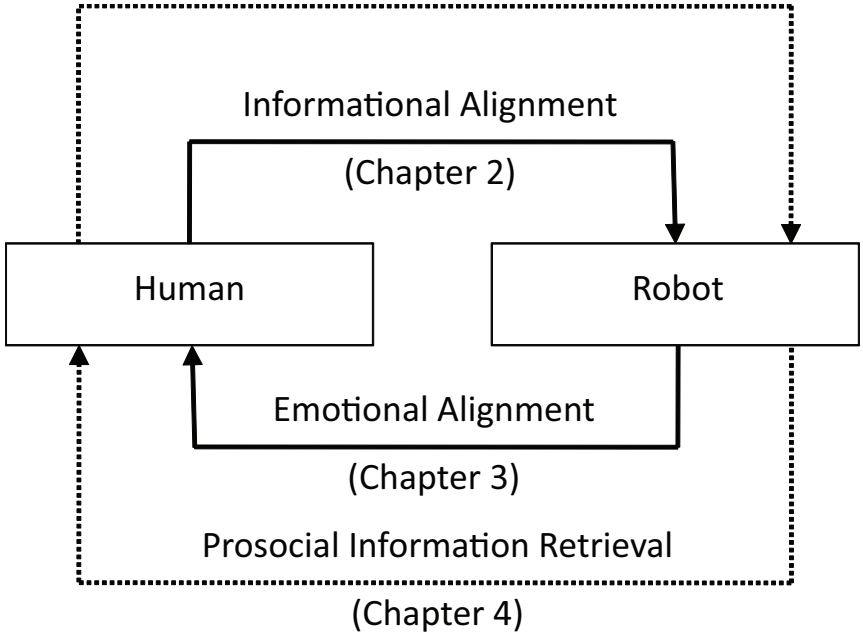

Figure 1.3: Outline of the thesis

The main contributions of this work are presented in the following.

Informational Alignment
The thesis targets robotic applications, where missing task-knowledge has to be retrieved from humans before the task is executed autonomously. Hence, a permanent informational alignment has to be conducted: firstly, between the robot and the human during information retrieval to assure that the information extracted by the robot meets the intentions of the human, and secondly, a re-evaluation of the extracted information during task-execution within the real world. Thus, it is necessary to represent and evaluate each newly gained information. A first contribution towards this goal is the development of a cognitive framework for information retrieval including not only HRI itself, but also information representation and real-world evaluation, as deduced from cognitive theories on human perception processes.

Regarding information retrieval from natural language HRI, this thesis investigates how informational alignment can be pro-actively controlled by the robot in a natural way for the user and with regard to varying speech recognition performance in outdoor environments. Thus, dialog strategies are developed to control the dialog structure dependent

on the environmental noise conditions. Thereby, suitable speech input is triggered, while providing a natural and intuitive interaction design to the user as far as the environmental conditions allow. Adequate handling of errors and miscommunication has to be ensured in order to provide validity of the acquired information on the one hand, and a satisfactory experience for the human interaction partners on the other hand. Along these lines, the main types of potential miscommunication are classified and assigned to corresponding states of information retrieval before being integrated in the dialog system.

Since speech recognition is a bottleneck for HRI in outdoor environments, two aspects have to be considered: While a highly structured dialog leads to improved speech recognition results, it is more unnatural than a less predefined dialog. In contrast, a more open dialog strategy makes predictions on how the answers of human conversation partners will look like severely more difficult, thereby making high speech recognition performances unlikely. In Chapter 2 these aspects of informational alignment are addressed in a systematical investigation of different dialog strategies with incorporated miscommunication handling requests. Finally, a switching mechanism is developed in order to adapt the dialog strategy to varying speech recognition conditions, incorporating smooth transitions from open to closed requests to maximize information extraction when needed, and to switch back to more open requests as soon as speech recognition recovers again.

Emotional Alignment

Emotional alignment is investigated in order to make use of emotional communication modalities to establish and maintain a prosocial interaction context. Thereby, the main contribution is a proactive control of the emotional alignment with humans by means of targeted prosocial biases deduced from social-psychological principles, integrated in a behavior control model for the robot.

In Chapter 3, emotional alignment is investigated in terms of triggering prosocial behavior towards a robot. In order to achieve this, in a first step the induction of situational empathy towards a robot is explored, triggered by different ways of emotional facial expressions animation, thereby revealing positive effects on subjective system-performance and other aspects of user experience (UX), and -acceptance. In a second step, a new methodological emotional adaption approach to trigger more prosocial human reactions in terms of increased helpfulness towards a robot is developed, deduced from social-psychological principles of human-human interaction, and enhanced by emotional prosody in speech. Unlike other state-of-the-art approaches, this approach proactively triggers a predefined target behavior for the task-benefit of a robot by transferring predictions on human behavior from social psychology to HRI, resulting in a behavior control model.

Since emotional alignment is highly associated with the legibility of emotional expressions, a comparative video-based online-survey is conducted in Chapter 4, before testing the generalizability of the approach embedded in an integrated architecture for prosocial information retrieval with the IURO-platform. Thereby, a systematical assessment of the design-dependent legibility of emotional expressions in speech and mimicry is conducted between two differently designed robotic heads of either machinelike versus more humanlike design. Potential differences in the user-perception on the HRI-key concepts of anthropomorphism and animacy are considered. An additional and more generalizable research question is to reveal potential impacts of dispositional empathy on the human performance in identifying the animated emotions.

Prosocial Information Retrieval

In order to fulfill a task that depends on information retrieval from prosocial HRI, a robot needs to manage social and cognitive sub-tasks. As an example, by triggering helpful behavior of humans, the robot is robust against dynamic environmental changes, which cannot be pre-programmed. Thereby, the request of the robot for help as well as the willingness of the human to help, can be regarded as social meta communication that serves as a motivational basis for information transfer, e.g., missing task knowledge. However, the task cannot be executed before the missing information is successfully extracted from natural language dialog. Thus, informational and emotional alignment have to be interlinked in relation to the task-success of the robot.

One contribution in Chapter 4, is a general underlying framework that integrates both, social and cognitive sub-tasks. Informational alignment is integrated in the cognitive subtask of information retrieval, interlinked with emotional alignment in the social sub-task of triggering helpfulness in human users in an implicit an explicit way of prosocial behavior control.

Another contribution is an integrated architecture for prosocial information retrieval, that is implemented in the dialog system of the robotic IURO-platform to be used in an evaluative outdoor application of the integrated approach. Thus, proactive information retrieval and prosocial behavior control are combined in form of a social sub-dialog prior to the route inquiry-dialog: While implicit emotional adaption in terms of emotional facial mimicry is used to increase the social capabilities and trigger empathy towards the robot, a common ground of the interaction is created by the integration of small-talk before entering the task-related part of the interaction. Moreover, prosocial behavior control is extended by an emotion recognition module in order to align with humans in a fully automated way during information retrieval.

An outdoor field trial is conducted in Chapter 4 to evaluate the integrated approach of prosocial information retrieval. The experimental results indicate that pro-active information retrieval in outdoor environments benefits from being combined with emotional adaption, since it reinforces the underlying prosocial motivation of humans to help a robot, thereby even compensating decreases in user acceptance due to bad speech recognition performance.

The aspects addressed in this thesis contribute to a fundamental understanding of prosocial information retrieval as an integrated concept. Although, a variety of information retrieving systems exist, only few methodical approaches are known exploiting their particular nature. In conclusion, this work contributes with 1) triggering suitable task-relevant information input from human users while compensating for poor speech recognition performance in outdoor environments, 2) insights on the transferability of social-psychological principles to HRI and deduced behavior control mechanisms 3) developing, evaluating, and applying an integrated framework of alignment strategies for prosocial information retrieval from natural language HRI. This work is highly interdisciplinary by applying human interaction patterns from linguistics and social psychology to robotic applications. The thesis is assigned to the context of information retrieval from natural language human-robot interaction (HRI), yet especially the socio-cognitive character of the framework for task-related HRI, dividing a task into cognitive and social sub-tasks, find use in a wider range of applications in task-related HRI.

It is the aim of this work to bring together and integrate very different and interdisciplinary facets of informational and emotional alignment in order to act as a guidepost and source of inspiration for future research in this field.

2 Proactive Retrieval of Missing Task-Information from Natural Language

This chapter is concerned with the investigation of informational alignment strategies for proactive information retrieval from humans in order to use this information source to obtain missing task-information, see Figure 2.1.

In order to achieve this goal, in a first step, a theoretical framework for proactive information retrieval (IR) from natural language is developed. On the one hand, relevant cognitive theories concerning human perception serve as a conceptual basis for the framework. On the other hand, the framework is deduced from findings about human-human communication patterns and coping strategies for miscommunication. The novel approach is, firstly, to combine these communication patterns with coping strategies and cognitive theories from human-human interaction (HHI) and, secondly, to transfer them to HRI as a general framework for proactive IR and handling of miscommunication. More precisely, natural error handling is achieved by selective raising of informational contents by means of well-directed requests at such a rate that miscommunication can be compensated. The presented approach is applicable to any task-oriented dialog. Given that asking for directions is a challenging example for task-oriented dialog between humans and a robot, the conversational context is exemplarily confined to route descriptions in public spaces.

Since the communication quality in public spaces is often impaired by noisy environment, it is difficult for a robot to retrieve missing task-information from humans. Thus, in this chapter, different dialog strategies are modeled and evaluated with respect to user experience and error handling capabilities in order to cope with erroneous speech recognition. Since correct recognition of spoken language is a bottleneck for real-world dialog systems, special emphasis is placed on the issue of adapting dialog strategies to the conditions under which the dialog is held to thereby provide for adaptability of the dialog strategy to variable speech recognition performance. Experimental evaluations are conducted in a fully automated indoor setting, and in a Wizard-of-Oz (WOz) outdoor setting. In consideration of the indications deduced from the experimental evaluations of the approach, an on-line switching mechanism is developed, implemented in form of an on-line switching dialog strategy, and experimentally evaluated.

The remainder of this chapter is organized as follows: A problem description is given in Section 2.1. In Section 2.2, a framework for proactive IR is developed, deduced from cognitive- and social-psychological theories, transferred to HRI. In Section 2.3 four different dialog strategies are modeled based on human communication patterns using smooth transitions between open and closed requests in order to adapt to varying environmental noise-conditions while maintaining best possible naturalness for the user and the raise of informational contents for the robot. Further, the dialog strategies are enriched by targeted miscommunication handling requests. The approach is evaluated in two different experiments in Section 2.4. According to the indications deduced from the experiments, an automated online-switching dialog strategy is developed and evaluated with regard to its applicability in varying background-noise conditions in Section 2.5.

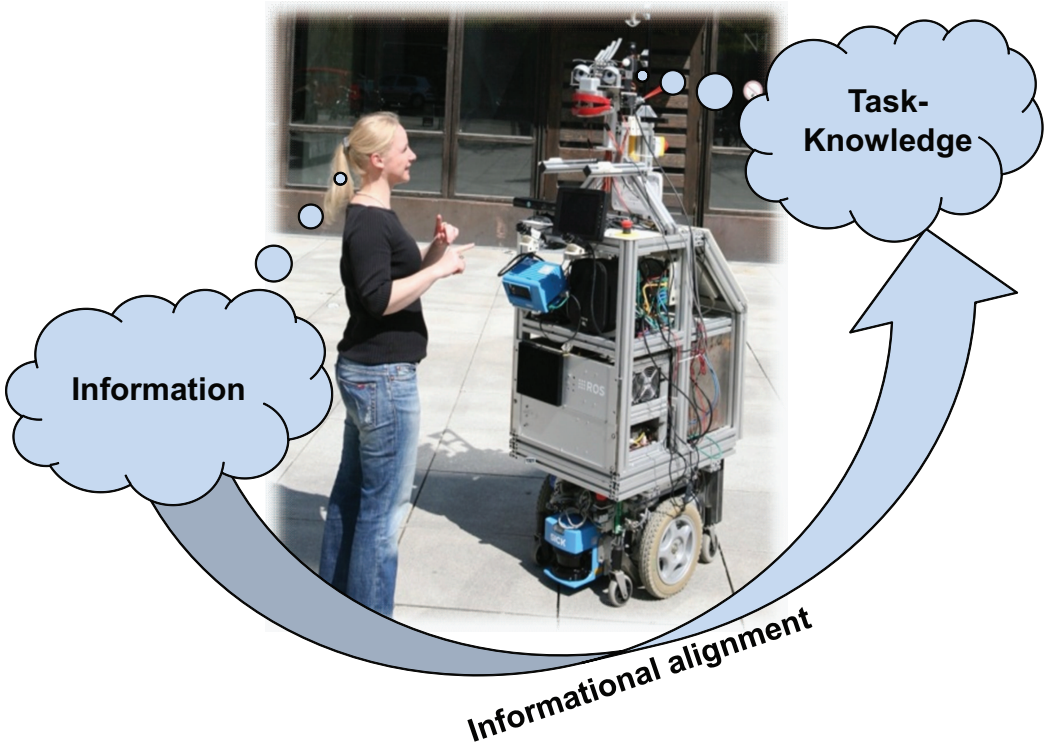

Figure 2.1: Pro-active retrieval of missing task-information from humans

2.1 Problem Description & State of the Art

In Human-Robot Interaction (HRI), the retrieval of knowledge from humans is vital in situations where a robot cannot fulfill its task based on the information it has or can acquire from its own sensor information. For robots in unknown and changing environments, this is a common situation. In order to have this information transfer work reliably, a robust dialog system is necessary, able to detect and handle errors and miscommunication while taking into account the task-relatedness of these dialogs.

Generally, natural language is the modality of choice for relaying task-related information to technical systems if easy accessibility and naturalness of the interaction are required and a training of possible users is not wanted or possible. For the information retrieval task in the urban setting, the use of natural language dialog is also justified according to the modality selection formulated by Kulyukin [87], since on the one hand the robot is autonomous in large portions of its behavior (navigation, action selection etc.), but on the other hand it also depends on the information retrieved from humans for fulfilling its task. Since robots are not restricted to humanlike communication modalities, in previous work also non-humanlike feedback modalities are explored by Mirnig et al. [58, 103, 104] in order to assess their use in conveying the internal system status to the user in information retrieval dialogs. The results show that non-humanlike feedback modalities are a good supplement to humanlike communication modalities by raising the efficiency of information retrieval as long as they are readable for human users, such as a screen to reassure understanding by re-visualized depictions of the information extracted from natural language. However, a clear trend towards verbal utterances being the most important modality to convey the internal system status to human users is approved.

However, the use of natural language for HRI also entails a number of difficulties, e.g. vagueness and ambiguity of spoken language itself, and the technical challenges posed by automatic speech recognition. As discontinuation due to weak speech recognition performance prevents any information retrieval from humans in public spaces, proper handling of errors and miscommunication has to be ensured in order to provide validity of the acquired information on the one hand, and a satisfactory experience for the human interaction partners on the other hand. Most existing approaches, where robots ask humans for missing task-knowledge in order to extract information about the environment are still operating in very simple structured indoor environments. For example, coarse qualitative route descriptions can be given to a wheelchair robot [106] that navigates in an office floor. The office robot Jijo-2 [2] learns the locations of offices and staff by moving around and asking humans for information. A robot asking for the way at a robotics conference is presented in [102]. A miniature robot that can find its way in a model town by asking for directions is described in [89]. These robots are able to interpret and follow simple route instructions, but cannot cope with the complexity and vagueness of natural language. Thus, careful design and robustness of the dialog is required, as well as adequate environment modeling for the situatedness of the dialog. Since speech recognition in outdoor environments is highly apt to be incorrect, resulting non- or misrecognition may eventually lead the dialog astray. Current approaches either use open requests, e.g. "how may I help you", and then classify the recognized speech input by means of machine learning in a second step [62], [35], or the other way is to use a dialog strategy, where the systems asks rather closed questions and thus breaks the task down into several subtasks in order to get more and more required information with every question [21]. Again, these approaches show the need for dialog strategies for a robot in order to control the dialog structure and thus triggering suitable speech input on the one hand, and to provide a sense of naturalness and intuitiveness for the user on the other hand.

As there is no control over the environmental conditions, which may have great influence on speech recognition performance, non-/misrecognition can occur frequently and eventually mislead the dialog. Hence, miscommunication has to be handled. Several studies successfully explored miscommunication and informational complexity arising from users giving verbal route instructions in a simulated dialogic real-time interaction with a robot executing the route instructions during the experiment [82, 135]. In contrast to state of the art approaches, this work addresses a robot that executes previously gained route instructions autonomously within real environments. Thus, complexity and the range of potential errors increase enormously, e.g. informational misalignment may be undetected during the dialog but lead to errors during execution of the gained route knowledge. Thus, it is necessary to represent and evaluate the every new information in a cognitive framework for information retrieval including not only HRI, but also information representation, and real-world evaluation. Apart from background noise, miscommunication may be caused by lexical or conceptual difficulties. Hence, another important aspect is to identify and differentiate between several potential types of miscommunication. Successful information retrieval and naturalness can be balanced by applying targeted handling requests in a flexible way, but integrated in a dialog strategy.

In the following section, a cognitive framework for proactive information retrieval from humans is developed, motivated by human perception processes and states of understanding.

Table 2.1: The process of perception in three consecutive stages according to TPH, adopted from Lilli & Frey [93]

Stage 1	Provision of expectation hypotheses
Stage 2	Input of information about the object of perception
Stage 3	Confirmation (end)/ disproof (restart) of the hypothesis

2.2 Framework for Proactive Information Retrieval from Humans

This section is concerned with the development of a cognitive framework for proactive information retrieval (IR) from humans, allowing for hypothesis-driven top-down information processing, but also considers bottom-up evaluation of the received stimulus input from natural language by incorporating consecutive states of understanding deduced from human perception processes, transferred to HRI in order to be used in spoken dialog systems (SDS). Previous developmental stages of the framework are published in [28, 57, 61].

2.2.1 Background from Social Psychology

Cognitive theories in social psychology are deduced from empirically proven data concerning human behavior and problem solving. Thus, they provide useful guidelines to be considered within HRI.

The *Theory of Perceptual Hypotheses* (TPH), as originally formulated by Bruner & Postman [27] and extended by Lilli & Frey [93] as *Hypotheses Theory of Social Perception*, is based on regarding perception as a cognitive interaction between on organism and its environment. Thereby, the process of perception is generally formulated as reception and interpretation of stimuli managed by available hypotheses about the environment. The basic assumption is that any process of perception starts with an expectation hypothesis, even before recognition of any environmental stimulus input. Such hypotheses originate from prior experiences of perception and can be seen as a set of cognitive predispositions. Accordingly, the chosen hypothesis affects perception to a certain degree by defining the kind of information to look for. Hence, the perceived objects can be seen as a selection out of diverse environmental stimuli organized by emphasizing some aspects of stimuli more than others. In other words, every perception can be seen as a result of former perceptions, successfully approved in prior similar situations. According to TPH, the process of perception consists of three consecutive stages as shown in Table 2.1, starting with the selection of an expectation hypothesis about the following environmental information input. Subsequently to the information input the process either ends with the confirmation of the selected expectation hypothesis, or, if the received input data does not match the hypothesis, the perception process restarts with the selection of a new expectation hypothesis. If this cycle restarts several times the underlying strategy may be falsification of several hypotheses. Hence, perception can be seen as a decision process.

The strength of Hypotheses

As there can be more than one hypothesis at the same time the actual extent of influencing perception performance depends on the strength of a hypothesis. If an underlying hypothesis is very strong it primarily determines the perception process, i.e. concept-driven or top-down information processing. In contrast, if the built hypothesis is rather weak it leads to data-driven information processing, i.e. bottom-up [167].

The strength of a hypothesis depends on five determinates, formulated by Lilli & Frey [93]:

1. The frequency of former confirmation: the higher the frequency the higher is the subjective confidence.

2. The number of alternative hypotheses: the higher the amount of alternative hypothesis, the lower is the chance for each hypothesis to take effect.

3. Motivational impacts: generally motivation triggers selection of hypothesis-supporting stimulus information and avoiding hypothesis-contradicting information.

4. Cognitive impacts: The more a hypothesis is fixed within cognition, the more it is dominant and modification-resistant, forming a guiding hypothesis e.g. daily routines.

5. Social impacts: In the absence of suitable stimulus information the accordance of social group members can serve as hypothesis confirmation.

There is a continuous relation between the strength of an expectation hypothesis and the available stimulus information input [93]:

1. The stronger a hypothesis, the more likely it is activated, i.e. *priming*.

2. The stronger a hypothesis, the smaller the amount of needed information to confirm it.

3. The stronger a hypothesis, the larger the amount of needed conflicting information to disprove it.

As a result, TPH provides a framework for evaluating the gained knowledge via hypothesis-testing and thus detecting and handling miscommunication for HRI, suitable for every task-oriented dialog.

States of Understanding

Empirical HHI-studies revealed that, within verbal communication, successful understanding evolves from the listener's ability to pass through four consecutive phases, called the "states of understanding" by Clark & Schaefer [36], depicted in Table 2.2. Miscommunication can affect each state, and parts of the same utterance may be spread over different states of understanding. In case the interpreter supposes to be in a more advanced state than she really is, the communicative goal is not achieved until a mutual belief about being in final State 3 is established for both interlocutors, where B understood what A meant by an utterance *u'*.

Table 2.2: Four consecutive States of Understanding by Clark & Schaefer [36]: B and A stand for listener and speaker, *u'* stands for any utterance of A.

State 0	B did not notice that A uttered any *u'*.
State 1	B noticed that A uttered some *u'* (but was not in state 2).
State 2	B correctly heard *u'* (but was not in state 3).
State 3	B understood what A meant by *u'*.

The states of understanding are the minimum basis for each successful communication act as passing them successfully assures that the listener understood what the speaker meant by an utterance. Thus, passing these states is indispensable for a robot talking to a human. Hence, these states provide a guideline for IR from humans, starting with speech detection. As miscommunication can occur in each state it is important to assign each category of miscommunication and related compensation strategy to one of these states. Thereby, a robot is enabled to choose the right compensation strategy for each kind of miscommunication depending from the state in which it occurs. The states of understanding are transferred to HRI as states of IR, and are integrated in the cognitive framework for proactive IR, as described more detailed in the following subsection.

2.2.2 Transfer to HRI: Framework for Proactive Information Retrieval

As introduced in [28, 57, 61], a dialog framework is developed in order to provide an all-embracing structure for proactive information retrieval in human-robot dialog, see Figure 2.2.

The Theory of Perceptual Hypotheses (TPH), as originally formulated by Bruner & Postman [27] and extended by Lilli & Frey [93] as *Hypotheses Theory of Social Perception*, provides the conceptual basis revealing that for humans any (mis-)interpretation results from a perceptual decision process evaluating an environmental stimulus input. This decision process can be seen as a loop controlled decision process composed of three stages: 1) provision of expectation hypothesis, 2) information input, and 3) confirmation or disproof of the hypothesis. In case of a disproved hypothesis, the cycle restarts as often as a hypothesis is confirmed. Transferred to spoken language dialog in HRI this means:

Stage 1) Expectation Hypothesis: Based on a context model the robot creates a hypothesis on what to expect from the human speech input, e.g. landmarks and directions.

Stage 2) States of Information Retrieval: In this stage, information input from HRI is requested. Thereby, a robot has to pass four states of IR, represented in a spoken dialog system (SDS) in order to extract the needed task-information from the speech input: *Speech Detection*, *Speech Recognition*, *Language Understanding*, and *Knowledge Representation*. As verbal miscommunication has to be handled in this stage, the robot is equipped with predefined associations between these states and different categories of miscommunication. For each of these, specific handling strategies are deduced from human-human corpora related to the corresponding states, see Figure 2.2. If all states of IR could successfully be passed the extracted conceptual knowledge is represented, e.g. in a route graph after each interaction.

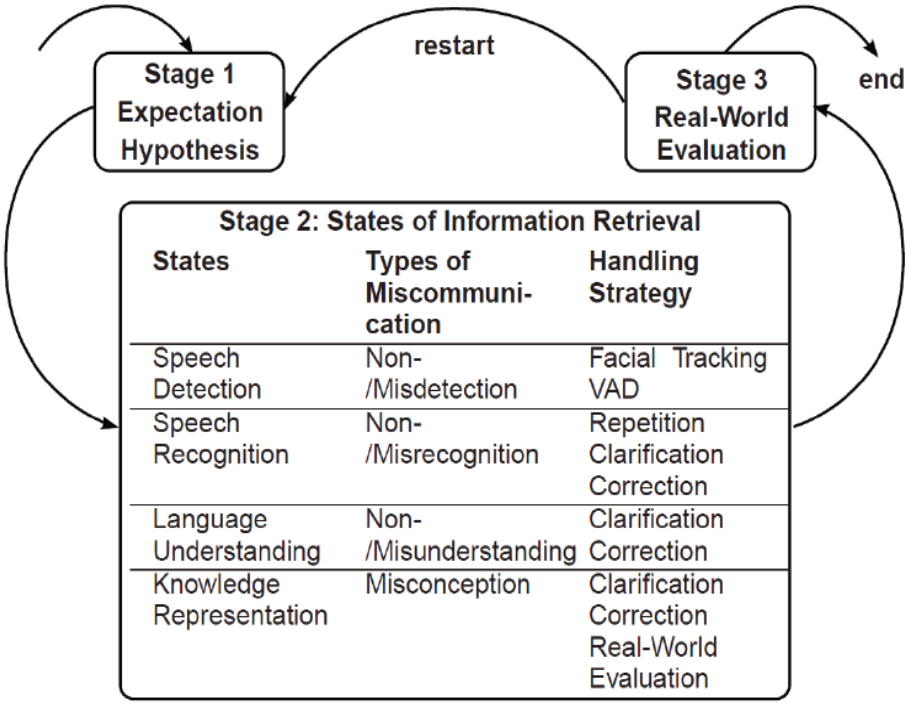

Figure 2.2: Framework for proactive information retrieval in HRI [28, 57]

In order to become aware of possible miscommunication, each state calculates a confidence value that indicates the extent in which the extracted information meets a hypothesis. In case of low confidence, e.g. caused by ambiguity in natural language, one of the related handling strategies is triggered. By employing handling strategies, the robot reduces the number of possible hypotheses regarding the interpretation of the speech input and thus raises the confidence values for other hypotheses until it is able to decide for one interpretation.

Stage 3) Real-World Evaluation: As miscommunication may be undetected during HRI, the robot has to evaluate the extracted information while performing its task. Therefore, the robot looks selectively for confirming or disproving information during task execution within the real world and, if necessary, confirms the information again by targeted questions on the desired task-status, e.g. if a certain landmark is actually reached or not. In case of different and/ or conflicting hypotheses from previous dialogs showing a common denominator, the task is performed until the critical point of conflict is reached, before the cycle of evaluation restarts with further hypothesis-testing in order to eliminate conflicting hypotheses, and decide for a new guiding hypothesis to be conducted and evaluated in the real world, see Figure 2.3.

After this overview of the functionalities embedded within the cognitive framework of proactive IR, the following two sections provide more detailed information on the design of **Stage 2)**, i.e. how information input from HRI can be arranged and actually carried out by the robot.

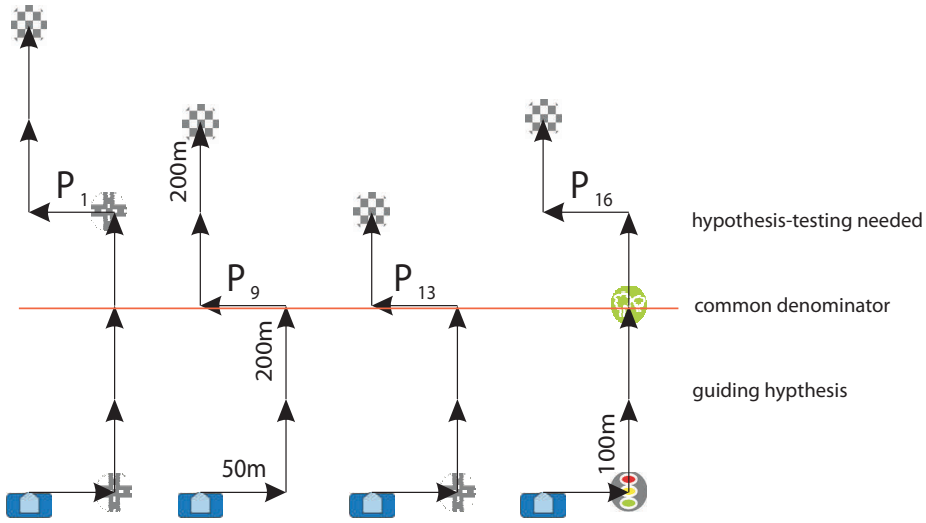

Figure 2.3: Four examples of route graphs, extracted from previous experiments [12, 13]. The red line marks the common denominator of the hypotheses as a critical point: below this line the route graph is confirmed and thus, turned into a guiding hypothesis. Above the line further hypothesis-testing is needed

2.3 Dialog Strategies & Miscommunication Handling Requests

In this section, four different dialog strategies are presented to be incorporated in **Stage 2)** of the cognitive framework of proactive IR. They allow adaption to inaccurate and unstable automatic speech recognition resulting from dynamically changing environmental impacts. Nevertheless, miscommunication may occur in each state of the information retrieval dialog. Thus, different handling requests are formulated for each state of the dialog in accordance with the corresponding category of miscommunication. The development of the dialog strategies and miscommunication handling requests is also published in [57, 58, 61].

2.3.1 Dialog Strategies

One approach to model human-robot dialog is to gather empirical data in a first step, e.g. by instructing humans to give directions to a robot without any verbal feedback while driving around in a building. In a second step, a conceptual route graph can be deduced that serves as a basis for later route inquiry dialogs [132]. When designing such dialogs, current approaches apply open requests, e.g. “How may I help you?”, and then classify the recognized speech input by means of machine learning [35, 62]. Another way is to use a dialog strategy, where the system asks targeted closed questions, e.g. ”Should I head in this direction?” and thus break the task down into several subtasks in order to selectively increase task-specific information with every inquiry [21].

In contrast to indoor settings, a robot retrieving missing task-knowledge from HRI outdoors, has to cope with unstable speech recognition performance depending on location-specific impact factors, e.g. noisy street vs. quiet park scenario. Hence, the approach is to integrate different dialog strategies incorporating smooth transitions between open and

closed requests, in order to adapt to good, fair, or bad speech recognition performance and thus raise the efficiency of information retrieval while maintaining naturalness of the interaction. Accordingly, the challenge is to develop different dialog strategies in order to enable the robot to adapt to varying environmental conditions for speech recognition.

In linguistic pragmatics, a common structure of four consecutive phases is identified by analyzing asking-for-directions dialogs [169], see Figure 2.4.

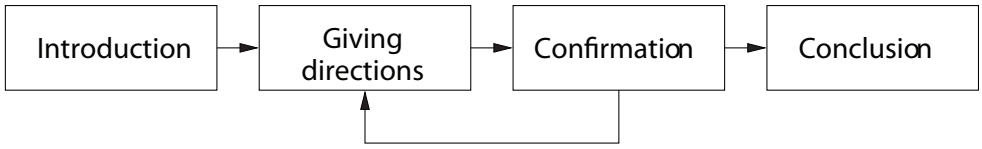

Figure 2.4: Common structure of asking-for-directions dialog [169]

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

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

Confirmation: Either of the two partners confirms the information. In this phase further inquiries can be made.

Conclusion: The asker thanks the respondent and they part.

This schematic structure is flexible, i.e. some phases may be interchanged or recur. Nevertheless, it is a well-proven guideline reflecting the intrinsic cognitive processes involved in human-human interaction and thus serves as a common ground to be transferred to HRI.

In the following, different dialog strategies based on the above-mentioned basic structure, but with variations regarding open, closed or mixed prompts, are presented. Over the different strategies, restrictiveness is gradually increased to gain more structured and thus more predictable dialog behavior. The strategies are exemplarily confined to the context of asking for directions but are applicable to any task-oriented inquiry dialog.

Open Dialog: This strategy exactly applies the basic structure of human inquiry dialogs and thus should be most intuitive for humans: The robot opens the dialog in *Introduction* phase by introducing itself and asking for the way to a certain goal location. After the human passer-by has given route instructions during *Giving directions* phase, the robot initializes the *Confirmation* phase by asking if it may repeat the entire route and subsequently asks if it was correct. If the reproduced route is not declared as correct by the human the robot requests to give the directions again and switches back to *Giving directions* phase to be repeated. After either the human interactant confirms the reconstruction given by the robot during *Confirmation* phase or refuses to repeat the instructions again, the robot thanks the human and thereby closes the dialog according to *Conclusion* phase.

Divided Dialog: The strategy coincides with Open Dialog regarding the *Introduction* phase. Yet, in *Giving directions* phase the robot asks directly for separate route segments by proactively opening this phase with the utterance "Please describe the first route segment". Subsequent to each explained route segment the robot asks if the route is already complete and requests the next route segment if necessary. During *Confirmation* phase

the robot repeats the route description by combining all obtained route segments, but asks for confirmation separately after reconstruction of each segment. Compared to Open Dialog, this strategy is designed to reduce the time spent by the human on correction through repeating only questionable route segments separately instead of repeating the whole instruction.

Requesting Divided Dialog: This strategy coincides with the structure of Divided Dialog but, unlike the latter, counts for each route segment in *Giving directions* phase if at least one landmark and one direction had been given or recognized. In case of no landmark, the robot requests the landmark by asking "How far should I go in that direction or up to which point?". Correspondingly, in case of no direction within a route segment, the robot asks "In which direction shall I head?" and afterwards inserts it into the reconstruction in *Confirmation* phase in order to be confirmed or corrected by the human interaction partner after each route segment. Just like in Divided Dialog, there is no reconstruction of the complete route at the end of the dialog to reduce the duration of the interaction.

Closed Dialog: This strategy differs from all other strategies regarding its flow: A user cannot give any free information input, but is asked to confirm or revise closed questions. Again, the robot introduces itself, but directly after asking for its way, the robot opens *Giving directions* phase and continues with closed questions like "Should I continue going in this direction?", "In which direction shall I head?" or "In which direction shall I turn then?", followed by "How far should I go in that direction" or "Up to which point?". Just like in the Requesting Divided Dialog strategy, the robot asks for directions and landmarks as long as it gets at least one of each for a route segment. Finally, it combines directions and landmarks to route segments in *Confirmation* phase in order to verify them by separated reconstruction. Accordingly, the human interlocutor has very limited input-possibilities, but speech recognition should be more robust due to the limited vocabulary.

As a conclusion, all dialog strategies incorporate the above-mentioned basic structure, but differ in *Giving directions* phase by allowing free speech input in Open Dialog, and turning more and more restrictive by requesting very concrete information in Divided and Requesting Divided Dialog until only closed questions are asked by the robot in Closed Dialog. Efficiency is varied in *Confirmation* phase with regard to route segments which can be confirmed or corrected separately in Divided-, Requesting Divided-, and Closed Dialog without the need of repeating the whole route as given in Open Dialog.

2.3.2 Miscommunication Handling Requests

In order to improve information retrieval within very noisy outdoor environments, Closed Dialog already contains requests to confine the vocabulary and to trigger the needed information input. Nevertheless, miscommunication may occur in all dialog strategies. Thus, there is additional need to assign targeted handling requests deduced from human-human corpora [52] to different categories of miscommunication [70].

In this subsection, the resulting types of miscommunication and corresponding handling requests are assigned to each state of the dialog, combinable with the above introduced dialog strategies.

In the proposed approach, route descriptions given by the subjects are stored and processed internally based on route graphs [166], representing a sequence of route segments. Each segment can consist of a *controller*, describing the traversal of a segment, a *router*

describing the location at the end of a segment, and an *action* to take once this location has been reached, e.g. a change of direction. This representation is used to form handling requests and to relate the described route to the user for possible correction. As the miscommunication handling requests are designed for experiments conducted in German language it is important to note that the following requests are translated as far as possible, but in some cases they meet the original meaning only approximately.

Repetition Requests: Assuming successful speech detection, miscommunication may initially occur in the state of speech recognition as "non-recognition", i.e. the robot could not gain any interpretation on what has been said by the human. In order to cope with non-recognition the following repetition requests are implemented:

"Could you repeat that, please?"/ "Excuse me, I didn't get your answer. + [Repetition of the previous question]"

Clarification Requests: As clarification requests are used to confirm an interpretation [52], these kinds of requests are employed in case of "mis-recognition". They are used as well in every case of miscommunication within all following states of information retrieval given that speech recognition already released one possible interpretation of the speech input which can be taken as a hypothesis in order to be confirmed by the following clarification requests.

Reprise sluices mark the interpretation gap by emphasizing "wh"-alliterated words:

"*Sorry,...where?/ when?/ how far?*"

Wh-substituted reprises repeat the well-understood part and substitute the interpretational gap:

"*Excuse me,...in which direction?/ up to which landmark?/ how far should I continue in this direction or up to where?/ in which direction should I turn then?/ how far should I go in this direction or up to where?*"

These particular requests are already integrated within *Closed Dialog* strategy as they are part of the closed questions in order to confine the vocabulary to facilitate speech recognition.

Reprise fragments are to emphasize an uncertain part of a gained interpretation:

"*Excuse me, did you mean...to the +[direction]?/ at/up to/near +[router]?/ I must pass by +[controller]?/ en route, I will see +[controller] on the right?*"

Alternative clarification questions are used to explicitly mention alternating interpretations in case of acoustic or referential ambiguity:

"*Excuse me, did you mean...sight or side?/ turn to the right or turn to the side?*"

"*Excuse me, ...*

...do I have to turn left or right at +[router]?" in order to confirm the direction.

...do I have to turn at + [router] or head straight on?" in order to confirm if a certain landmark depicts a *controller* and not a *router*.

...do I have to pass +[controller] or turn there?" in order to confirm if a certain landmark depicts a *router* and not a *controller*.

Task-level reformulations are used to clarify more complex actions by reformulating the consequences of an utterance and thereby demonstrating subjective understanding. Thus, these requests confirm the practical implication within an utterance:

"*This would mean that.....I have to turn back?/ I should not turn until I have passed +[controller]?/ at + [router] I have to turn + [direction]?*"

Correction Requests: If miscommunication is detected, e.g. by the user during *Confirmation* phase when the robot reconstructs the route description, correction requests are employed to revise an interpretation in order to determine the underlying intention of the speaker [52]:

"*Excuse me, I think I got you wrong,...please tell me where I have to go instead./ please give the directions again, and a bit slower.*"

The different dialog strategies are evaluated stand-alone, and in combination with the miscommunication handling requests in two different experiments, as described in the following section.

2.4 Experimental Evaluation

In order to evaluate the dialog strategies and miscommunication handling requests, motivated and developed in Section 2.3, the route description domain was chosen for the experiments as it provides a valid and rather well-explored structure for the extraction of missing task-knowledge from natural language HRI. The route descriptions given by the experimental subjects are stored and processed internally based on route graphs [166], representing a route as a sequence of route segments. As introduced in the previous section, each segment can consist of a *controller*, describing the traversal of the segment, a *router* describing the location at the end of the segment, and an *action* to take once this location has been reached, e.g. a change of direction. This representation is used to formulate miscommunication handling requests and to reformulate the described route to the user for possible correction during *Confirmation* phase.

The different dialog strategies are each modeled as state sequences according to their specifications given in Section 2.3. Each state could either be a textual output node with speech output generated from templates filled with stored information given by the user, and input nodes where information given by the user is entered into the internal knowledge representation of the system. The transitions between the nodes, determining the course of the dialog, are specified with conditions on the previous course of the dialog, e.g. requirements on the information given as response, such as posing a more refined question when not all relevant information had been provided.

The evaluation experiments are partially published in [57, 58].

2.4.1 Experiment I: Fully-Automated Indoor Setting

A first evaluation of user acceptance and user experience for the dialog strategies described in Section 2.3 is conducted in a user experiment. The dialog strategies, modeled as state sequences as described above, are used as templates for the DialogOS [1] tool, that provided text-to-speech synthesis and speech recognition for a fully automated (FA) natural language dialog. In this experiment, finite-state grammars for speech recognition are created for each input node.

[1] CLT Sprachtechnologie GmbH, www.clt-st.de

During the experiment, a route displayed on a map was presented to the subjects for each of the dialog strategies separately. Then, the subjects were asked to interact with the dialog system in order to describe the depicted route. Following each dialog, the subjects filled in a questionnaire by rating several items describing their impression of the dialog. The experiment was held in a laboratory setting without robotic embodiment of the dialog system.

2.4.2 Experiment II: WOz-Outdoor Setting

A Wizard-of-Oz (WOz) experiment is a commonly used method in the field of human-computer interaction with the goal of observing the use and effectiveness of a proposed user interface. In this kind of experiment the subjects interact with an apparently autonomous computer system but that is actually being operated completely or in parts by an unseen human being (the "wizard") [78]. The dialog strategies are evaluated in combination with the miscommunication handling requests in an outdoor environment, again in terms of user experience. In order to simulate good speech recognition performance, the task of entering user input into the knowledge representation of the system and the choice of system actions were performed by a human operator, that the subjects were unaware of.

For this, the wizard was asked to perform in a way similar to a system restricted in its input by a predefined grammar for the user answers depending on the dialog state, similar to the system used for the experiment described in Section 2.4.1. The transitions between the nodes were chosen depending on the dialog state by the wizard within the bounds of the respective dialog strategy, as defined by the corresponding state machine models. The MARY text-to-speech tool [130] was used to generate German synthesized speech according to the templates specified in the dialog models.

In addition to the mere modeling of succession of dialog states, miscommunication handling strategies were implemented in this experimental setting. After each input node in which relevant, task-related information had been requested, the wizard had the possibility to choose from a number of handling requests as described in Section 2.3.

In this experiment, for each dialog strategy, the subjects engage in dialogs with a dialog system embodied in a robot platform with human-like features. A map was not presented to the participants, but the users were asked to describe a way of their own choice to a well-known, nearby location.

In order to provide a realistic setting for the dialog, the experiment was conducted in an outdoor environment at Technische Universität München, and test persons faced the current state of the *IURO* robotic platform based on the Autonomous City Explorer (ACE) robot [13] as interaction partner. The interaction scenario is depicted in Figure 2.5. In order to enable natural language dialog and other modalities of interaction, the robotic platform is equipped with a number of sensors including cameras and microphones, a loudspeaker and a mechanical actuated head capable of displaying emotions and lip movements synchronized to speech [97].

2.4.3 Experimental Measures

After each interaction, subsequently to one of the four dialog strategies, which were presented in random order, the subjects of both experimental settings (**FA** and **WOz**) filled

Figure 2.5: Experimental setting for the WOz-experiment: Interaction with the robot, controlled by a wizard from inside the building.

in a questionnaire designed for measuring the user experience of the dialog strategies. In detail, these items are (*translated into English*):

comprehension: "The system understood what I said/IURO understood me well."

duration: "The duration of the interaction was appropriate."

expectation: "I always knew which comments the system/IURO expected from me."

structure: "The structure of the dialog was sensible."

correction: "When there was a misunderstanding, the correction effort was appropriate."

request: "The system/IURO asked wisely for missing or uncertain information."

satisfaction: "Overall, I was satisfied with the dialog."

Each item was rated by participants on a 5-item Likert scale ranging from *1 = "strongly disagree"* to *5 = "strongly agree"*.

2.4.4 Experimental Results

Results can be deduced from the initial fully automatic (FA) experiment including 16 subjects (11 male and 5 female, between 22 to 35 years with a mean of 27.0 years) described in Section 2.4.1 and the Wizard-of-Oz (WOz) experiment including 29 subjects (21 male and 8 female, between 19 and 39 years with a mean of 22.9 years) described in Section 2.4.2.

Regarding the reliability of the developed questionnaire on user experience, the coefficients of internal consistency for for the items are good (Cronbach's $\alpha > .80$ overall).

The significance level for all performed tests is $\alpha = .05$ except for multiple testing where it has to be adjusted using the correction method of Bonferroni.

Table 2.3: Mean ratings with standard deviations (in brackets) of single items and total scores for each dialog strategy derived from the Fully Automatic (FA) experiment (rated on Likert scales from 1 = strongly disagree to 5 = strongly agree).

	Dialog			
	Open	Divided	Requesting Divided	Closed
comprehension	2.9(1.3)	3.4(1.0)	3.2(1.1)	4.1(0.8)
duration	3.1(1.4)	3.3(1.0)	2.9(1.1)	3.8(0.9)
expectation	3.4(1.0)	4.1(0.9)	3.3(1.2)	3.7(1.2)
structure	3.6(1.4)	3.9(0.9)	4.0(0.9)	4.1(0.9)
correction	2.9(1.4)	3.5(1.2)	3.4(1.1)	4.1(1.0)
request	2.5(1.5)	3.6(1.2)	3.5(1.2)	3.9(1.1)
satisfaction	2.7(1.4)	3.3(0.8)	3.1(0.9)	3.9(1.0)
total score	3.1(1.1)	3.6(0.8)	3.3(0.7)	3.9(0.8)

Fully Automatic (FA):
According to the results of Kolmogorov-Smirnov tests, normal distribution could be accepted for every single item as well as for the total scores (calculated as means of all single item values per dialog strategy). Therefore, parametric comparisons and correlations are performed.

An analysis of variance (ANOVA) with repeated measures revealed no significant difference between the total scores of the four dialog strategies. However, mean values show a trend towards a difference between the total rating of the *Open Dialog* and the *Closed Dialog*. Further repeated measure ANOVAs analyzing the ratings of single items provided significant differences between the dialog strategies for *comprehension* ($F = 4.16, p = .011$), *correction* ($F = 3.72, p = .023$), *request* ($F = 7.89, p = .001$) and *satisfaction* ($F = 4.11, p = .012$). Post-hoc t-tests revealed a significant difference between *Open Dialog* and *Closed Dialog* for *comprehension* ($t = -3.31, p = .005$). Similar deviations indicating that *Closed Dialog* was rated higher than *Open Dialog* were also obtained for *correction*, *request* and *satisfaction*, but failed to reach significance due to Bonferroni correction of significance level to $\alpha = .0083$. Means of single item ratings and total scores of the different dialog strategies are displayed in Table 2.3.

Correlation analysis focused on the item *satisfaction* led to the finding of most meaningful relations in the *Open Dialog* condition: The general satisfaction of the subjects with the dialog significantly correlated with *comprehension* ($r = .83, p < .001$), *duration* ($r = .75, p = .001$), *structure* ($r = .79, p < .001$), *correction* ($r = .89, p < .001$) and *request* ($r = .98, p < .001$). The impression arises that more aspects of the interaction with the system had to be pleasing to satisfy the user in *Open Dialog* condition in comparison to the other strategies.

Wizard-of-Oz (WOz):
For the single items, normal distribution had to be rejected according to the results of Kolmogorov-Smirnov test, but was accepted for the total scores (again calculated as means of all single item values per dialog strategy). Hence, comparisons and correlations regarding

Table 2.4: Mean ratings with standard deviations (in brackets) of single items and total scores (rated on Likert scales from 1 = strongly disagree to 5 = strongly agree) plus number of handling requests for each dialog strategy derived from the Wizard-of-Oz (WOz) experiment.

Means of single items and total scores.

	Dialog			
	Open	Divided	Requesting Divided	Closed
comprehension	3.6(1.3)	3.9(1.1)	3.7(1.2)	4.0(1.0)
duration	4.2(1.0)	3.8(1.1)	3.5(1.3)	3.2(1.3)
expectation	4.0(1.0)	4.0(1.1)	4.0(1.0)	3.5(1.4)
structure	4.4(1.0)	4.3(1.2)	4.0(1.1)	3.9(1.2)
correction	3.9(1.2)	4.0(1.1)	3.7(1.2)	3.7(1.2)
request	3.8(1.4)	4.4(0.8)	4.2(0.9)	4.0(1.0)
satisfaction	4.3(0.8)	4.1(1.1)	3.9(1.2)	3.8(0.9)
total score	4.0(0.7)	4.1(0.8)	3.8(0.9)	3.7(0.8)

Number of handling requests per type and in total.

repetition	4	4	3	9
clarification	17	53	63	70
correction	5	1	3	7
total	26	58	69	86

the single items were performed non-parametrically and parametric methods were used for the total scores.

Again, no significant difference between the total scores could be derived from ANOVA with repeated measures, but means show a trend towards higher ratings of the *Open Dialog* and *Divided Dialog* compared to the remaining two strategies. On single item level, the ratings of *duration* ($\chi^2 = 9.00, p = .027$) and *satisfaction* ($\chi^2 = 8.62, p = .035$) significantly varied. According to post-hoc analyses the rating of *duration* considerably differed between *Open Dialog* and *Closed Dialog* ($Z = -2.68, p = .006$), whereas after adjusting the α-value no significance remained for *satisfaction*. Means of single item ratings and total scores of the different dialog strategies are displayed for all three conditions in Table 2.4.

Correlations of *satisfaction* with other single items again varied between the dialog strategies. In condition *Closed Dialog*, there were clear relations to *comprehension* ($r = .57, p = .001$), *duration* ($r = .59, p = .001$), *structure* ($r = .59, p = .001$), *correction* ($r = .57, p = .002$) and *request* ($r = .56, p = .002$). Fewer significant correlations could be obtained for the other strategies and there was only one in the *Open Dialog* condition (*comprehension*: $r = .51, p = .006$). Apparently, more aspects of the interaction are important to satisfy participants in the *Closed Dialog* condition compared to the other strategies.

In addition, to solve miscommunication problems, several handling requests could be used by the wizard during each dialog. The distribution of *repetition*, *clarification* and

correction requests was examined with Friedman tests. A significantly different use between the dialog strategies was only obtained for the total amount ($\chi^2 = 45.90, p < .001$) and for *clarification* requests ($\chi^2 = 41.76, p < .001$). These results mainly arose from deviations between *Open Dialog* compared to the other strategies. The absolute number of applied handling requests was the highest for *Closed Dialog* for every type of request and in total. The number of included requests, both for different types and for all requests in sum per dialog, is displayed in Table 2.4.

Comparison of ratings between both experiments:
Total scores of strategy ratings were compared between the **FA** and **WOz** experiment with paired t tests resulting in a significant difference for *Open Dialog* ($t = -3.09, p = .005$) and marginally significant deviations for *Divided Dialog* ($t = -1.98, p = .054$) and *Requesting Divided Dialog* ($t = -2.00, p = .052$). The means indicate higher ratings for *Open Dialog*, *Divided Dialog* and *Requesting Divided Dialog* in the **WOz** compared to the **FA** experiment, whereas for *Closed Dialog* the relation is vice versa. Means and standard deviations of total scores for both experiments are displayed in Figure 2.4.4. In sum, quantitative results show varying ratings of the different dialog strategies between the two conducted experiments. In **FA** *Closed Dialog* was rated highest and *Open Dialog* lowest, whereas in **WOz** *Closed Dialog* was the most unpopular strategy. Surprisingly, in the second experiment most handling requests were used in *Closed Dialog* and least in *Open Dialog*.

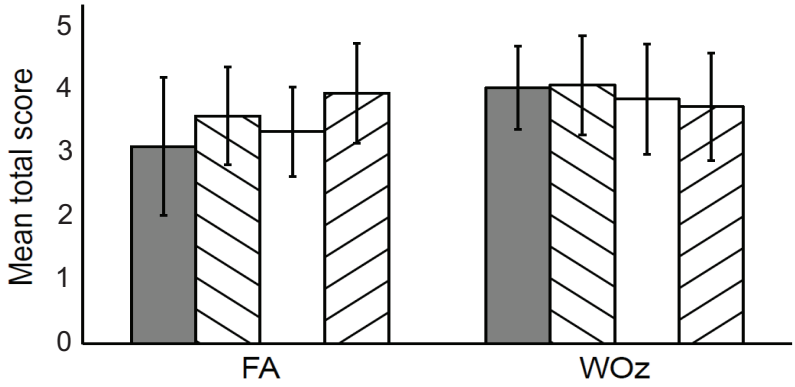

Figure 2.6: Means and error bars (± one standard deviation) for the total scores both of the Fully Automatic (FA) and the Wizard-of-Oz (WOz) scenario for all different dialog strategies. From left to right, the bars for each condition describe ratings obtained using *Open*, *Divided*, *Requesting Divided* (yellow) and *Closed Dialog* strategies, respectively.

Duration of interaction:
Actual durations of the interactions per dialog strategy are distributed normal in each condition and experiment. Hence, comparisons are performed parametrically.

Concerning the interaction durations of the conditions between both experiments, paired t-tests revealed significant results for *Divided Dialog* ($t = -2.74, p = .009$), *Requesting Divided Dialog* ($t = -3.29, p = .002$) and *Closed Dialog* ($t = -2.75, p = .009$). As indicated by the means, interactions in these conditions were clearly longer in the **WOz** compared to the **FA** experiment (means and standard deviations of durations per dialog strategy for both experiments are displayed in Figure 2.4.4). Longer durations in these condi-

tions probably resulted from the high amount of included handling requests as depicted in Table 2.4.

One significant difference was obtained within the **WOz** experiment ($F = 11.13, p < .001$). According to post-hoc tests, only the actual duration of *Open Dialog* strongly deviated from all other conditions. This finding fits the results derived from the questionnaire, in which the perceived duration of the *Open Dialog* strategy was rated best. For the initial **FA** experiment, no significant differences between the four strategy-durations were found and again, this fact suits the duration ratings derived from the questionnaire.

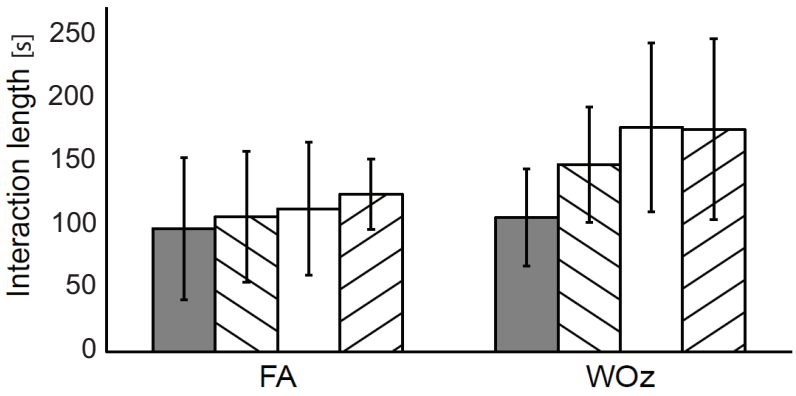

Figure 2.7: Means and error bars ($\pm$ one standard deviation) for the duration of dialogs both of the Fully Automatic (FA) and the Wizard-of-Oz (WOz) scenario for all different dialog strategies. From left to right, the bars for each condition describe ratings obtained using *Open*, *Divided*, *Requesting Divided* and *Closed Dialog* strategies, respectively.

2.4.5 Discussion

The fact that the total user experience scores within *Closed Dialog* condition decrease marginally in **WOz** compared to the previous **FA** experiment goes hand in hand with the usage of handling requests which is highest within *Closed Dialog* and lowest within *Open Dialog* condition. This finding might speak against the acceptance of the proposed handling requests. However, in all other conditions the total scores are raised compared to the **FA** experiment without handling requests, which indicates a slightly positive impact of those. Additionally, the above mentioned decreasing trend within *Closed Dialog* might be due to the fact that the most used *clarification* requests were already partly implemented in this condition, because of being part of the closed dialog strategy. This might have caused a feeling of over-usage of handling requests for the users compared to the other conditions. Furthermore, as the wizard was only able to employ handling requests while being the initiative part within all dialogs, *Closed Dialog* provided more chances for usage than within *Open Dialog*, because there was least turn-taking due to the open prompt-strategy at the beginning of the dialog which did not allow the wizard to interrupt the user while giving route instructions. In contrast, the *Closed Dialog* strategy provokes frequent turn-taking and thus allows for more handling requests.

Indications

Summing all up, the experimental results indicate that the application of handling requests raises user satisfaction as can be seen in the total scores for *Open*, *Divided*, and *Requesting Divided* Dialog. However, at a certain point, when there are too many handling requests employed, the effect changes into the opposite and decreases user ratings again. Another factor of influence is the duration of the interaction: Within the **WOz** experiment the duration of *Open Dialog* condition is significantly shorter than all other dialog strategies and accordingly rated as most convenient duration in the questionnaire. Due to the resulting significant increase of the total scores for *Open Dialog* compared between **FA** and **WOz** experiments it is suggested to employ the proposed handling strategies in a flexible way, but confined in a way to avoid a critical increase of dialog duration and a feeling of over-usage.

Thus, in the following Section, the implementation and evaluation of an online-switching dialog strategy is described in order to adapt to varying speech recognition performance in outdoor environments.

2.5 Handling of Varying Speech Recognition Performance

In the previous Sections 2.3 and 2.4, four different dialog strategies for a robot asking human passers-by for directions are modeled and evaluated in two different experiments [58, 60]: a fully automated indoor experiment, and its replication in an outdoor Wizard-of-Oz (WOz) setting, where all types of requests for miscommunication handling are added and evaluated in combination with each dialog strategy. According to the experimental results, the quality of conversation between robots and humans highly depends on the performance of speech recognition: Poor speech recognition performance not only drastically decreases the amount of retrieved information for the robot but also naturalness of the interaction for the user. On the one hand, the performance can be greatly enhanced by employing a more closed dialog strategy, while open dialogs tend to be more intuitive to human users.

Since speech recognition is a bottleneck for HRI in outdoor environments, two aspects have to be considered: While a highly structured dialog leads to improved speech recognition results, they are more unnatural than a less predefined dialog. More open dialog strategies, on the other hand, make predictions on how the answers human conversation partners will look like severely more difficult, thereby making high speech recognition performances unlikely.

Thus, in this section an approach is developed to improve information retrieval in outdoor environments with varying speech recognition performance while maintaining highest possible naturalness of the interaction for the human interaction partner. The benefit of the online-switching dialog strategy with integrated miscommunication handling requests for non- or misrecognized task-knowledge is twofold: On the one hand, the gain of information is stabilized by the possibility to switch from open to closed requests with reduced grammar needed, and thus, improved speech recognition performance in case of very noisy environmental interaction conditions. On the other hand, naturalness of the interaction is maintained in two ways: 1) by avoiding an over-usage of miscommunication handling requests following the open questions through switching to closed questions if the background noise level changes during an interaction, 2) by providing the possibility to switch back to a more open dialog strategy as soon as informational alignment with a human interaction

partner has recovered, e.g., due to decreasing background noise. Critical thresholds in background noise are evaluated in this section.

2.5.1 Online-Switching Dialog Strategy

In order to prevent non-/misrecognition of the information input, a switching mechanism is developed that adapts the dialog strategy to the quality of speech recognition. An algorithm monitors the informational alignment during an interaction by calculating an online-confidence score that triggers switching to a more closed dialog as soon as the informational alignment decreases under a certain threshold during an interaction. If informational alignment recovers again, e.g., due to a reduced environmental background noise level, the algorithm triggers the transition back to a more open dialog in order to maintain highest possible naturalness for the user by employing open questions to retrieve the missing task-knowledge. Additionally, in order to handle potentially occurring non-/misrecognition, miscommunication handling requests (MHRs) are integrated.

In the IURO-Project, the robot operation system (ROS)[2] is used to manage the communication between the different modules of the robot. Thus, the modules of dialog management and speech recognition are connected, with all dialog-components being represented and accessible by ROS calls and services. The online-switching dialog strategy is implemented and evaluated in the IURO-dialog system based on IrisTK: a statechart-based toolkit for multi-party face-to-face interaction [136]. The dialog structure takes the form of a finite state machine (FSM) implemented through XML, which is then compiled into Java-Code and connected with the ROS-architecture by a Python script. The FSM can handle external input as well as Java-scripting for added features.

The online-switching dialog strategy is based on two previously developed dialog strategies, as introduced in Section 2.3, selected to be combined by the switching mechanism: Since the noise-conditions in an urban outdoor environment are not expected to allow for a completely open dialog, the **Requesting Divided Dialog** strategy is selected for better speech recognition conditions, and the **Closed Dialog** strategy is chosen for poor speech recognition performance. In order to handle potentially occurring non-/misrecognition, all three types of MHRs are integrated as described earlier in Section 2.3: **Repetition Requests**, **Clarification Requests**, i.e. ***Wh-substituted reprises***, and **Correction Requests**, using slots and fillers to clarify uncertain or missing parts of a route description. The detailed design of the selected dialog strategies is described in the following.

Requesting Divided Dialog

For good/fair environmental speech recognition conditions, this dialog strategy is designed as depicted in Figure 2.8, where components of the dialog-structure are depicted in rectangles, verbal output and input in rounded rectangles: After the *Introduction* phase, the phase of *Giving directions* is divided into route segments, that are initially requested by an open question, starting with "Could you please describe the first segment of my route?" in order to trigger suitable user-input. After the open answer of the user, a feature of the XML-architecture is used: The grammars, defining what the robot is listening for after asking an open question to the user allows the identification of directions, and landmarks or distance information in different lexical and grammatical variations. If at least one landmark and one direction could be identified, the route segment is preliminary regarded

[2] http://www.ros.org/, 2013

as successfully recognized. In case of total non-recognition (no landmark and no direction could be recognized), a **Repetition Request**, e.g., "Could you repeat that a bit louder, please?" is employed to get the open user-input again. In case of partial recognition (either landmark or direction is missing), **Clarification Requests**, in particular, ***Wh-substituted reprises*** are used to retrieve the missing landmark or direction by targeted closed requests, e.g., "In which direction shall I turn then?" and/or "How far should I go in that direction?", integrated in the dialog strategy. In case of non-recognition of the following closed answer, the system uses a **Repetition Request** again. In order to avoid a feeling of MHR-overusage, as indicated by the experimental results in Section 2.4, a dialog-abortion is implemented as soon as three MHRs have been asked without recognizing any answer. If the answer can be recognized the dialog proceeds with the *Confirmation* of the retrieved route segment by repeating the retrieved landmark(s) and direction(s) and asking the user, if they are correct. In this state of the dialog, misrecognition can be handled, e.g., if the dialog system misrecognized a "left" as a "right". If the human notices any misrecognition, the robot asks the same to specify and correct the error. If the error affects both, landmark(s) and direction(s), the system again parses the open user input.In case of either a misrecognized landmark or direction, the same ***Wh-substituted reprises*** are used again for clarification, as described above. In case of a correct route-repetition, the system goes on with the next route segment in an analog way, until the human interlocutor states the end of the directions, thereby triggering the dialog system to conclude the dialog in *Conclusion* phase.

Although the **Requesting Divided Dialog** strategy already contains a transition to closed requests in case of partial non-recognition or misrecognition of a route segment, it would always restart the following route segment with an open question. In order to avoid open questions at the expense of dialog-efficiency, **Closed Dialog** is applied in case of high background noise.

Closed Dialog

For poor environmental speech recognition conditions, this strategy is implemented in collaboration with Skantze [3], as can be seen in Figure 2.9. In this strategy, non-recognition of open user-input is assumed due to very noisy environmental impacts. Thus, directly after introducing itself and it's task, the robot opens the *Giving directions* phase with closed **Clarification Requests** like "Should I continue going in this direction?" or "In which direction shall I turn then?", followed by "How far should I go then?" or "Up to which point?" and, thereby, directly requests one direction and one landmark or distance information for each route segment, as long as the goal location is indicated by the user. For every non-recognized information, a **Repetition Request** can be asked. After each route segment, consisting of two closed **Clarification Requests** for one direction and one landmark, a *Confirmation* phase is conducted in order to handle miscommunication in case of misrecognized route information by retrieving the erroneous direction or landmark again with the the related **Clarification Request**. Accordingly, the human interlocutor has very limited input-possibilities, but speech recognition should be more robust due to the limited vocabulary.

In order to avoid endlessly requesting dialog loops, e.g., in very noisy environmental conditions or if a human interlocutor leaves the interaction without completing the route description, a dialog-abortion is again implemented as soon as three MHRs have been

[3] Ph.D., Department of Speech Music and Hearing, KTH Stockholm

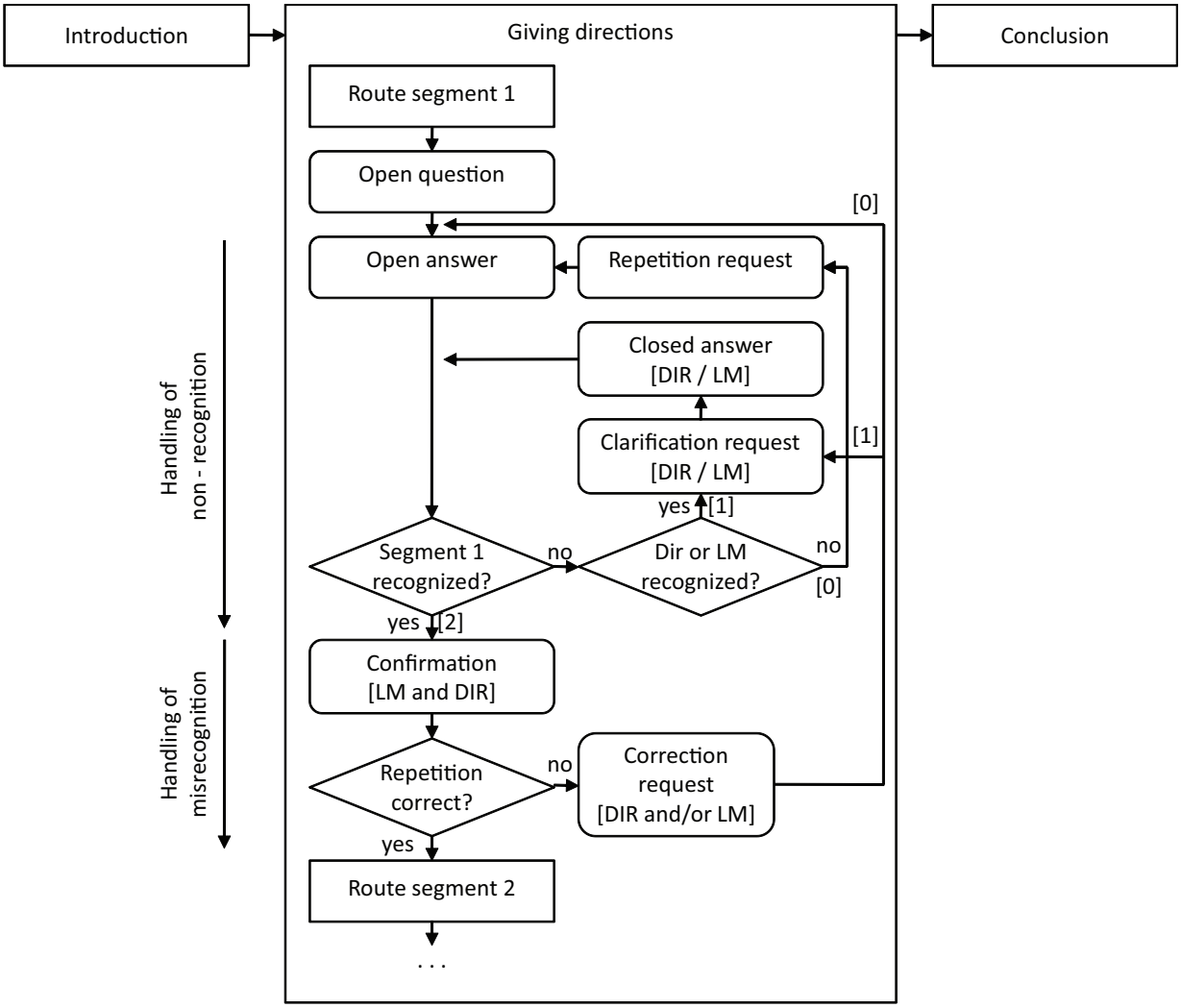

Figure 2.8: Dialog structure of the *Requesting Divided Dialog* strategy

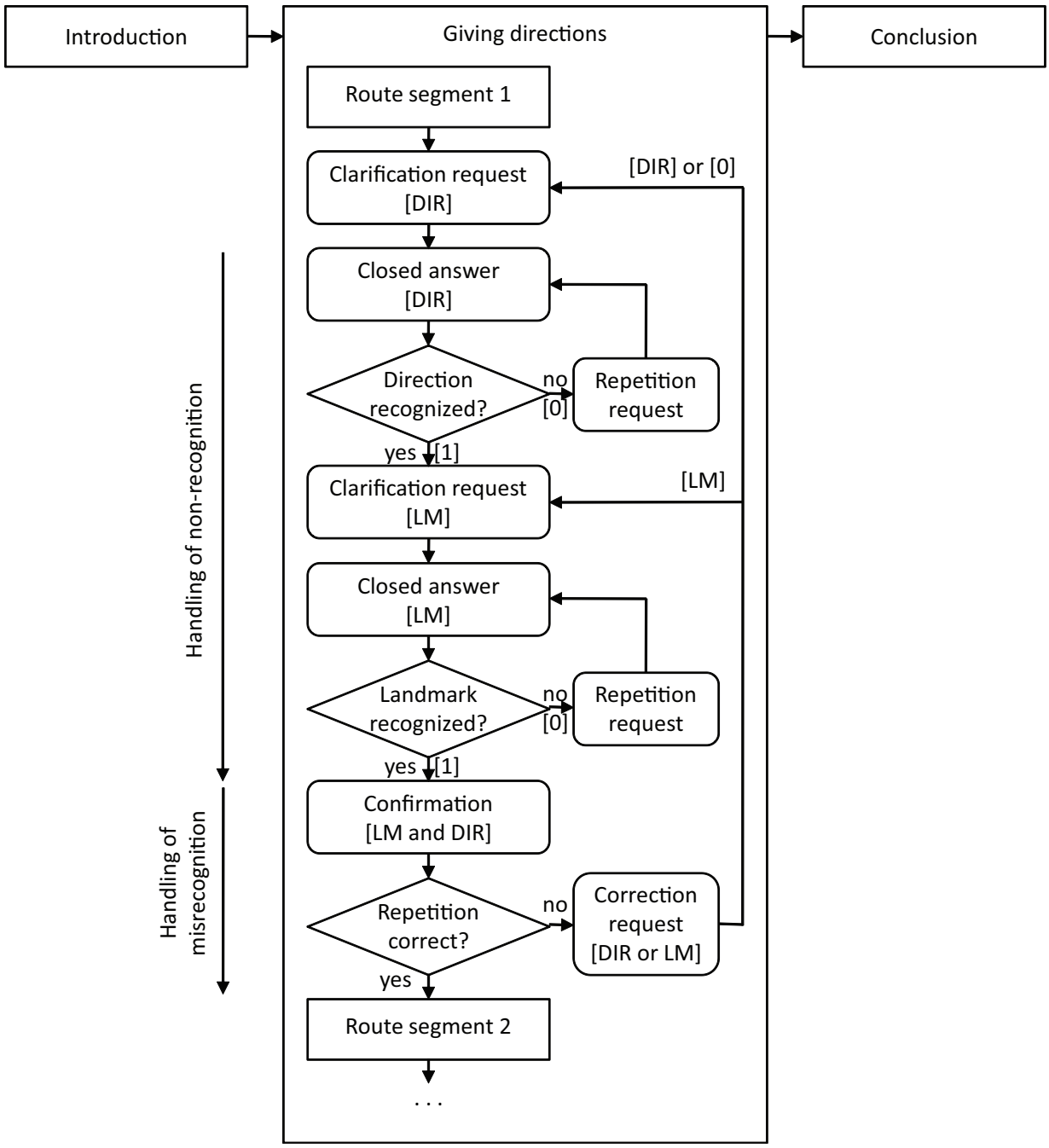

Figure 2.9: Dialog structure of the *Closed Dialog* strategy

asked by the robot without recognizing any answer. In order to provide the opportunity of switching back to open questions in terms of naturalness for the user if the speech recognition stabilizes in dependence of the noise level again, a switching mechanism is developed as described in the following.

Switching Mechanism

In order to prevent non-/misrecognition of the information input and to avoid an overusage of MHRs as far as possible, a switching mechanism is developed that adapts the dialog strategy to the quality of speech recognition. Therefore, the informational alignment during an interaction is monitored by calculating an online-confidence score in order to trigger switching between **Requesting Divided Dialog** and **Closed Dialog**. As soon as the informational alignment decreases under a certain threshold ε during an interaction, a switch to **Closed Dialog** is triggered. If the informational alignment recovers again, e.g., due to a reduced environmental background noise level, the algorithm triggers the transition back to **Requesting Divided Dialog** to maintain highest possible naturalness for the user by employing open questions to retrieve the missing task-knowledge.

The basic principle of the online-switching dialog strategy is depicted in Figure 2.10.

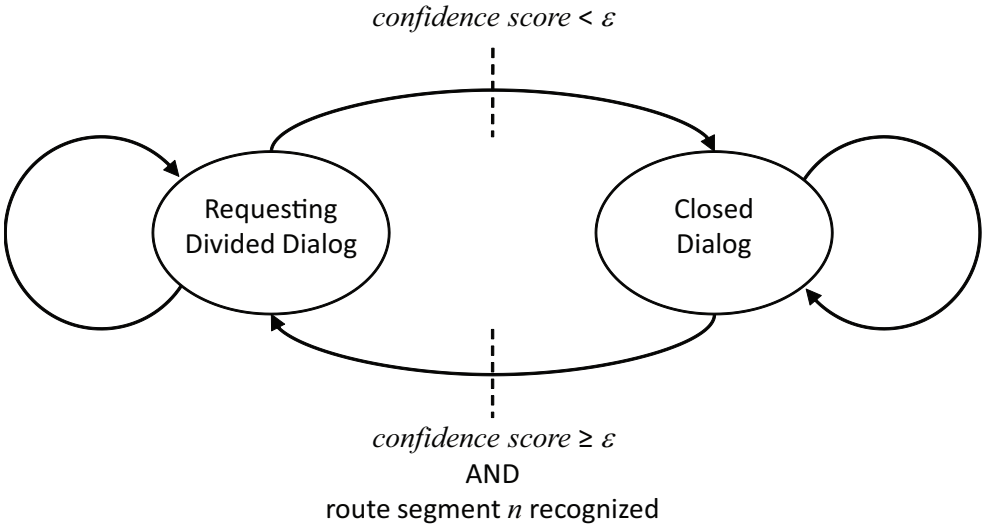

Figure 2.10: Basic principle of the online-switching dialog strategy

For the switching mechanism, the number of employed MHRs in case of non-/misrecognized information is monitored online during information retrieval: Whenever miscommunication occurs, i.e., a **Repetition-**, **Clarification-**, or a **Correction Request** is used by the system to clarify non-recognized information in the recognition check to retrieve a route segment, or misrecognized information in the *Confirmation* phase, the confidence is decreased, while each successfully recognized information improves the confidence. The system starts with a value just high enough to try the **Requesting Divided Dialog** when first approaching his conversation partner in the *Introduction* phase, switching to **Closed Dialog** if the confidence score decreases under the critical threshold ε, and back to **Requesting Divided Dialog** if the confidence score increases again and passes the threshold. Thereby, the timing of switching is additionally defined by the structure of the respective dialog strategy in use: Whereas a switch to **Closed Dialog** is possible during a route segment in **Requesting Divided Dialog**, the switch back to **Requesting**

Divided Dialog is not conducted until a route segment is finished in **Closed Dialog**, because once a route segment is started to be retrieved via closed **Clarification Requests** it would not make sense to ask the open question for the next route segment in **Requesting Divided Dialog**. Thus, switching back to the more open dialog strategy is started with the next route segment.

In summary, the rules for the switching algorithm look as follows:

Algorithm 3.1 Online-Switching Dialog Strategy

```
1:  confidence = ε
2:  miscom = 0
3:  do
4:      Requesting Divided Dialog
5:      if information entity retrieved then
6:          confidence = confidence +α
7:      endif
8:      if miscommunication
9:          confidence = confidence - α
10:         miscom = miscom + 1
11:     endif
12:     if confidence ≥ ε then
13:         Requesting Divided Dialog for next route segment
14:     else
15:         Closed Dialog
16:     endif
17: until interaction goal achieved or miscom ≥ δ
18: Conclusion
```

2.5.2 Experiment III: Evaluation of the Online-Switching Dialog Strategy

In order to evaluate the adaptability of the online-switching dialog strategy to different noise levels, an experiment is conducted for different noise levels, but with constant speech input. The noise level is simulated as an experimental variable in order to evaluate potential influences of environmental background noise. Uncontrolled additional variables like speaker-differences, lexical and/or acoustical interferences are avoided and kept constant in the experimental setup. The goal of the experiment is to evaluate the switching mechanism, and to reveal critical noise levels for the application of the online-switching dialog strategy in outdoor environments.

Experimental Design & Measures

In order to ensure experimental reproducibility, and to exclude lexical or acoustical interferences, it is important to keep the speech input as constant as possible. Thus, the speech input is restricted to one route description, consisting of a "yes" to simulate a positive answer to the request for help of the robot in *Introduction* phase, followed by two identical route segments consisting of one landmark and one direction each in *Giving Directions* phase. The route description ends with stating the goal by the utterance "arrived". Accordingly, six information entities have to be retrieved by the dialog system.

In order to avoid the uncontrolled influence of speaker-differences, the route instruction is previously recorded as an audio-file, to be presented to the dialog system as answers to the requests. Due to the absence of real users, the structure of the dialog is confined to the *Introduction* phase, followed by *Giving Directions* phase, without *Confirmation* and *Conclusion*. Hence, in this evaluation only total- and partial non-recognitions of the route segments can be considered by the confidence score since the algorithm adds α scores for each recognized information, even in case of misrecognition that can only be detected in *Confirmation* phase, where α scores are deduced when indicated by the user. In this evaluation, misrecognitions are post-hoc analyzed.

For the experimental setup, a quiet room is chosen to avoid uncontrolled noise interferences. The dialog system opens the conversation with a confidence score of $\varepsilon = 50$ before it asks for help and then adds $\alpha = +10$ scores for each recognized information, and subtracts $\alpha = -10$ scores for each non-recognized information, that has to be additionally requested. However, the dialog is aborted if the number of miscommunications (in this experiment non-recognitions) δ passes a threshold of 3. An audio-player is placed in front of a microphone to simulate the user-input in a constant distance to the same. In order to simulate noisy outdoor conditions for speech recognition, pink noise is selected corresponding to urban street noise [158] as expected in the IURO-project. The pink noise is generated with MATLAB[4] [138], measured with an error of +/- 3.5dB(C).

Five different noise levels are evaluated in five dialog runs, from 40 to 80 dB(C). As experimental measures, all runs are transcribed and dialog performance is evaluated by counts conducted for the number of:

- switchings to **Closed Dialog**
- re-switchings to **Requesting Divided Dialog**
- successfully recognized dialogs
- handled non-recognitions by **Repetition Requests** or by **Clarification Requests**, not being part of the respective dialog strategy
- not handled non-recognitions including dialog abortions and post-hoc analyzed misrecognitions

The experimental results are presented and discussed in the following.

Experimental Results

Results are deduced from 25 dialog runs, where five runs were conducted for each experimentally simulated noise-level condition: 40 dB(C), 50 dB(C), 60 dB(C), 70 dB(C), and 80 dB(C).

Table 2.5 shows the proportionate occurrences of the dialog performance-measures in five runs each for the noise-levels of 40 to 80 dB(C): "Successful dialogs" indicates the number of dialog runs (out of five), where all requested information entities could be successfully retrieved by the dialog system at the end of the dialog. "Dialog abortion" specifies the number of dialog runs, where the system aborted the dialog because three consecutively unsuccessful attempts to handle non-recognition. The number of dialog

[4] http://www.mathworks.de/products/matlab

Table 2.5: Proportionate occurrences of dialog performance-measures in five runs each for the noise-levels of 40 - 80 dB(C)

Dialog performance	Noise-level (C)				
	40dB	50dB	60dB	70dB	80dB
Successful dialogs	5	5	5	2	0
Dialog abortion	0	0	0	1	5
Occurrence of non-recognition	0	0	4	5	5
Handled non-recognition	0	0	4	2	0
Not handled non-/misrecognition	0	0	0	3	5
Switching to closed	0	0	3	3	5
Re-switching to divided until dialog-end	0	0	2	2	0
Confidence means	110	110	88	66	24

runs, where non-recognition occurred are indicated by the third measure in the table. "Handled non-recognition" counts the dialog-runs where all occurred non-recognition could be successfully handled, whereas "Not handled non-/misrecognition" includes all dialog runs that were either aborted because of three consecutively unsuccessful attempts of non-recognition handling (one run at 70 dB (C)), or misrecognized information that occurred in the second route segment of two different runs at 70 dB (C), respectively: Once a landmark, and once the indication of the direction was misrecognized. Since the *Confirmation* phase was omitted in the experiment, the two misrecognized information entities are post-hoc analyzed, and counted as not handled since it is not known if it would have been detected and corrected by a human user. The number of dialog runs are indicated by "Switching to closed", where at least one switch to **Closed Dialog** is conducted, and the row below indicates the dialog runs that could be completed in **Requesting Divided Dialog** due to a recovery of the confidence score that could be sustained until the end of the dialog run. Finally, the confidence means over five runs each, are indicated.

Figures 2.11 and 2.12 illustrate these proportionate distributions of the switches, and of the dialog performance-measures. As can be seen in both figures, the noise-levels most relevant to the online-switching dialog strategy are 60 and 70 dB(C): The occurrence of non-recognition starts at a noise level of 60 dB(C) where all occurring non-recognitions could be handled by the dialog system online. Whereas, the total amount of online-switches is the same for 60 and 70 dB(C), the handling of non-recognition failed in one run that led into dialog abortion, and two cases of misrecognition occurred at a noise level of 70 dB(C). With a noise level of 80 dB(C), the recognition of information in **Requesting Divided Dialog** did not exceed the initial "yes", as exemplarily depicted in Figure 2.13. As a consequence, all dialog runs of 80 dB(C) resulted in a switch to **Closed Dialog** (see Figure 2.11), where no further information could successfully be retrieved though.

In the left of Figure 2.13, an exemplary dialog transcript for 40 and 50 dB(C) is shown, where all information was recognized by the dialog system without any non-/misrecognition in all five runs each. In contrast, in the exemplarily selected transcript for 80 db(C) on the right, only the initial "yes" was recognized in two of five runs, while in the remaining

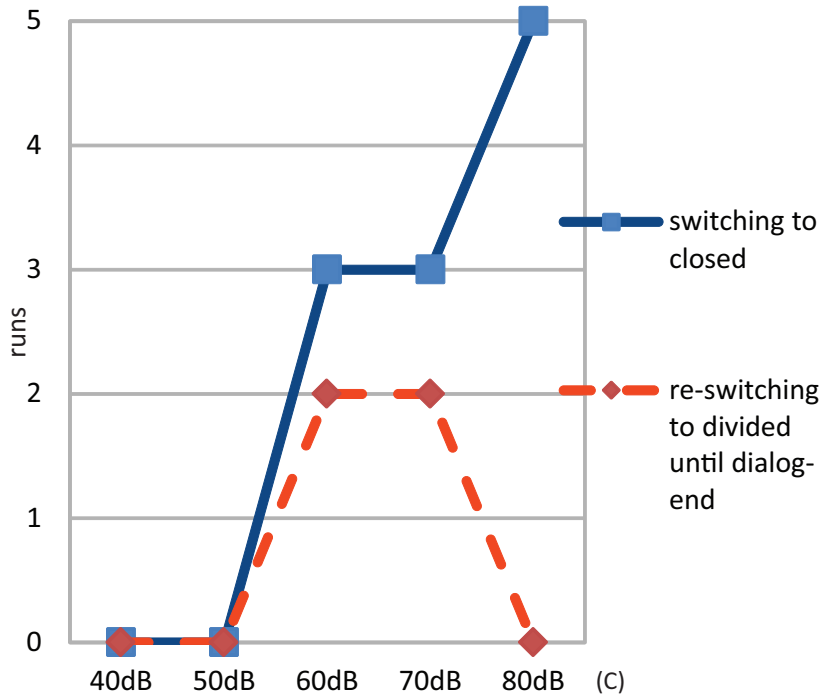

Figure 2.11: Proportionate distribution of online-switching in the dialog runs from 40dB(C) to 80dB(C)

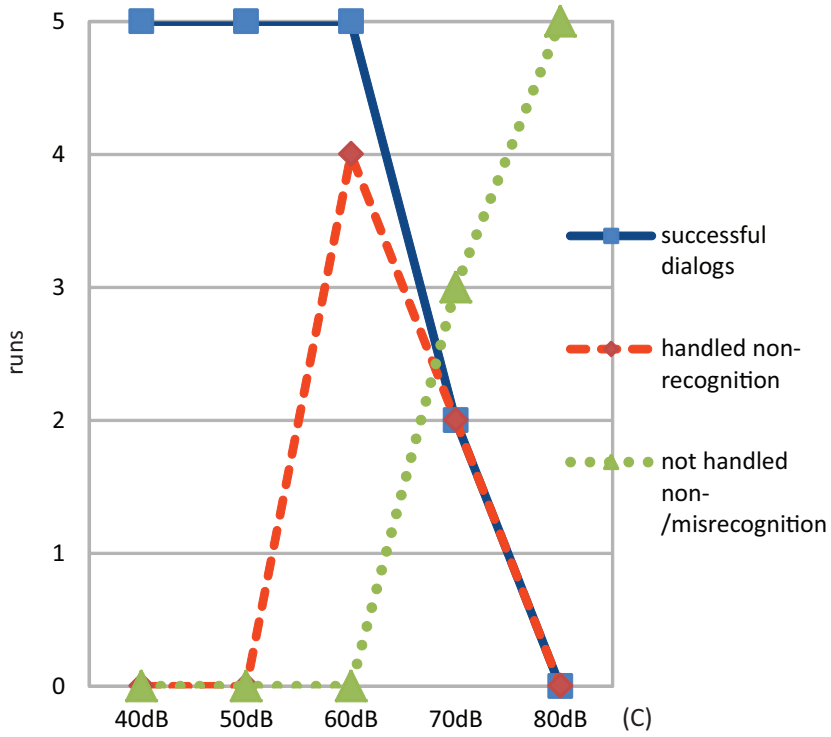

Figure 2.12: Proportionate distribution of dialog performance-measures in simulated noisy environment from 40dB(C) to 80dB(C)

three runs, all information was non-recognized. All transcripts show the development of the confidence score, recognition or non-recognition of the information entities from the speech input, and tentative online-switches between the **Requesting Divided Dialog**, referred to as "divided" and the **Closed Dialog**, referred to as "closed", corresponding to the dialog progress on the x-axis. For 80 dB(C), all five runs were aborted due to three consecutively unsuccessful MHRs.

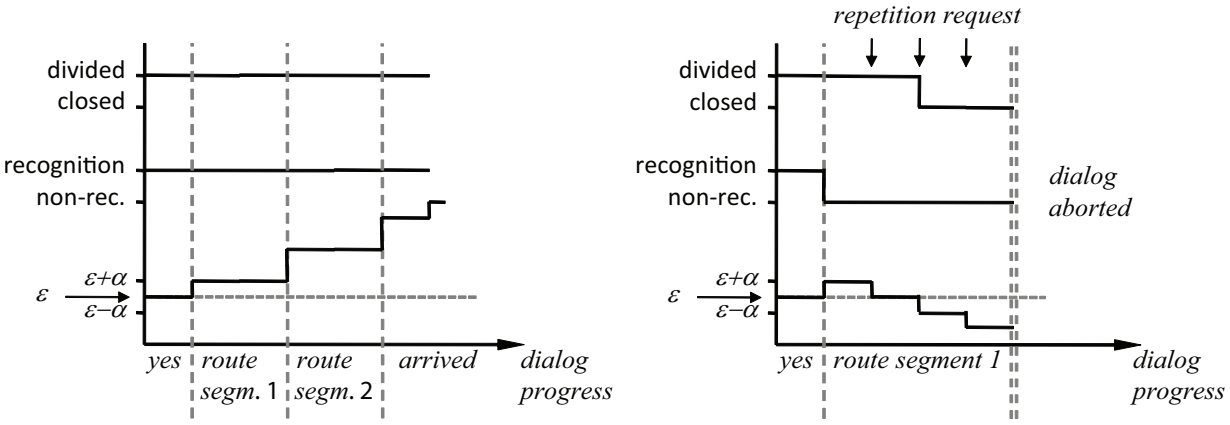

Figure 2.13: Transcribed online-switching dialog results for pink background noise of 40 to 50 dB(C) on the left, and for 80 dB(C) on the right

The total occurrences within these dialog runs resulted in 0 non-recognitions for 40 and 50 dB(C) accordingly, and 9 non-recognitions, requested by MHRs (8 **Repetition Requests** and 1 **Clarification Requests**) for 60 dB(C), which is an improvement compared to previously measured 11 non-recognition requests for 60 dB(C) using only **Requesting Divided Dialog** without the switching mechanism. The highest amount of 19 MHRs (13 **Repetition Requests** and 6 **Clarification Requests**) were employed for non-recognitions at 70 dB(C), followed by 15 MHR-requested non-recognitions at 80 dB(C), where the decrease of MHR-use (15 **Repetition Requests** and 0 **Clarification Requests**) is due to the dialog abortion in all five runs after three unsuccessful **Repetition Requests**. The total number of switches was identical for 40 and 50 dB(C) with no switches, and for 60 and 70 dB(C), each with 4 switches to **Closed Dialog** and 3 re-switches to **Requesting Divided Dialog**. At 80 dB(C), all five dialog runs switched to **Closed Dialog** after the non-recognized initial "yes" with no re-switches to **Requesting Divided Dialog** and thus, no confidence recovery, before the dialog abortion.

According to the experimental results, the noise levels most relevant to the online-switching dialog strategy is from 60 to 70 dB(C). In Figure 2.14, an exemplary transcript is depicted for 60 dB(C), where four switches were conducted, with two times to **Closed Dialog**, and back to **Requesting Divided Dialog**, as soon as the confidence score recovered to the threshold ε again in the first part of route segment 1, and at the other re-switch in route segment 2 until the end of the directions. Prior to the switches to **Closed Dialog**, two **Repetition Requests** were unsuccessfully employed by the system.

In Figure 2.15, an exemplary transcript for a dialog run at 70 dB(C) is shown, where two switches are conducted: The first switch to **Closed Dialog** is conducted after the non-recognition of the initial "yes", where the confidence score decreases under the threshold ε. The second switch is a re-switch to **Requesting Divided Dialog** with the confidence

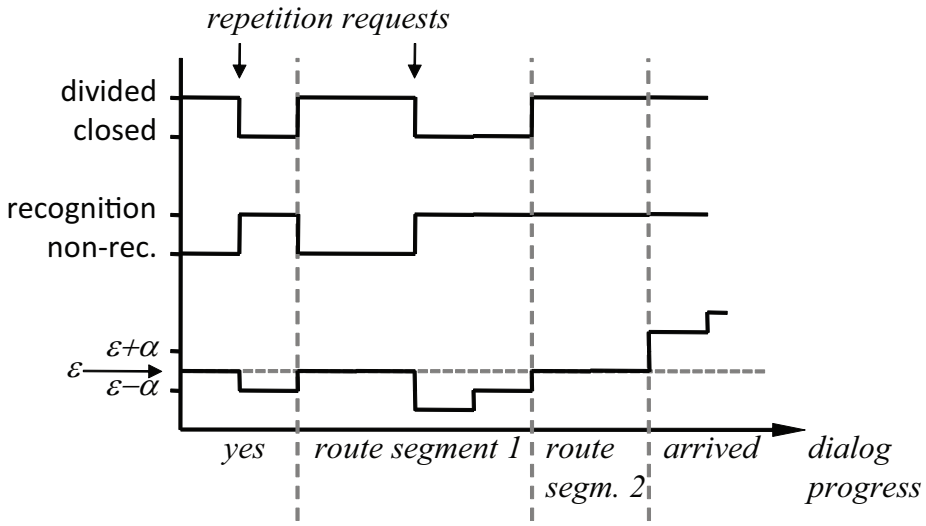

Figure 2.14: Transcribed online-switching dialog results for pink background noise of 60 dB(C)

score reaching the threshold again for route segment 1. In route segment 1, only one information entity ([DIR])was recognized at the first place, the second information entity ([LM]) had to be requested by a **Clarification Request**. In route segment 2, the first information entity ([DIR]) could be identified as misrecognized during post-hoc analysis, and thus, would have been to be handled in *Confirmation* phase, where $\alpha = -10$ scores would have been subtracted if the user detected and indicated the miscommunication. However, since a *Confirmation* phase was not part of the evaluation, the algorithm adds $\alpha = +10$ scores for a successfully recognized information entity that would be subtracted again in *Confirmation* phase. In this dialog run, the system employed three MHRs: one unsuccessful **Repetition Request** before switching to **Closed Dialog** for the retrieval of the first information entity "yes", one successfully employed **Clarification Request** for the landmark after retrieving only the first information entity ([DIR]) in route segment 1 being back in **Requesting Divided Dialog**, and finally another successful **Repetition Request** for the last information entity "arrived".

The experimental results and indications for the handling of varying speech recognition performance in noisy environments are summarized and discussed in the following.

Discussion

In the experimental evaluation of the online-switching dialog strategy, critical noise levels for the handling of varying speech recognition performance by the switching mechanism could be identified. In the experiments, there were no non-recognitions at the noise levels of 40 and 50 dB(C), resulting in successfully conducted information retrieval using **Requesting Divided Dialog** without using the switching mechanism. At 60 dB(C), non-recognitions occurred that could successfully be handled by the targeted application of MHRs combined with the switching mechanism to **Closed Dialog**, and back to **Requesting Divided Dialog**, where indicated by the confidence score, exceeding or falling below a threshold ε. At a noise level of 70 dB(C), the number of successfully handled non-recognitions started to decrease with first occurring dialog abortions after three consecutively unsuccessful MHRs, and miscommunication in form of misrecognized information entities emerged. This kind of miscommunication can potentially be handled in the

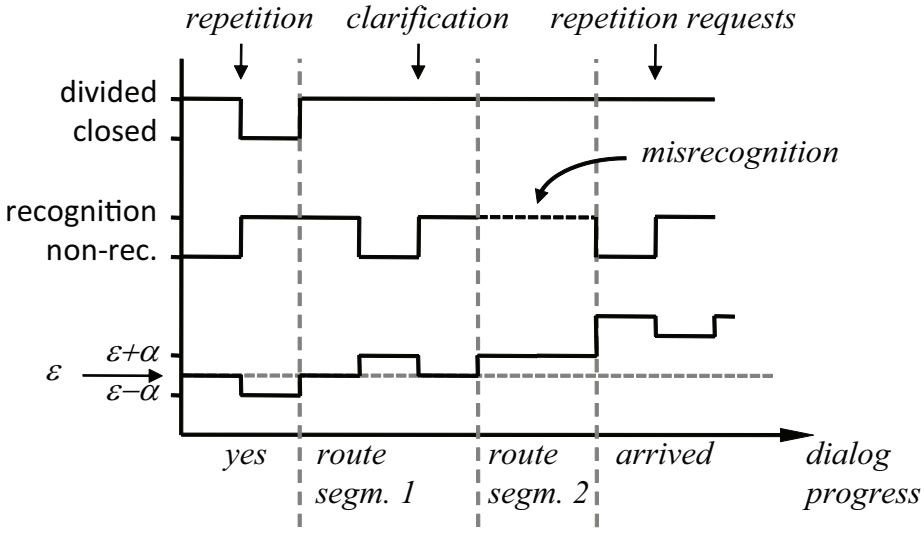

Figure 2.15: Transcribed online-switching dialog results for pink background noise of 70 dB(C)

Confirmation phase based on user-corrections, but can cause serious misleadings if they remain undetected by the user. Hence, 70 dB(C) is the critical noise level, where misrecognitions can lead the information retrieval in wrong paths, but that are still resolvable by **Correction Requests** in the *Confirmation* phase, included in both paths of the online-switching dialog strategy. However, at 80 dB(C), no successfully recognized dialogs could be achieved in the experiments. Not one non-recognition could be handled and despite switching to **Closed Dialog**, every run resulted in a dialog abortion after the application of three consecutively unsuccessful MHRs. In the presented experiments, the varying speech recognition performance, and corresponding confidence variations, that triggered the switching mechanism was due to stochastic nature of the noise signals within a noise level, and by fluctuations of the system performance itself. However, this is the more a positive indication for the applicability of the switching mechanism in outdoor environments with varying noise disturbances.

Indications

When interpreting the experimental results according to a table for noise-disturbance of "Umwelt-Bildungs-Zentrum" (UBZ)[5], the following indications are deduced: The online-switching dialog strategy is capable of handling non-recognitions due to varying speech recognition performance up to a noise level of 60dB(C), that is comparable to disturbing conversations up to a distance of one meter next to the robot. Due to the integrated *Confirmation* phase, the system is also able to handle potential misrecognitions at a noise level of up to 70 db(C), corresponding to the noise disturbances at a crowded place. Thus, the online-switching dialog strategy expands the spectrum of successful information extraction from 50 dB(C) up to 70 dB(C) by the application of a switching mechanism, leading the dialog in a closed dialog strategy where needed, and switching back to a more open dialog strategy in case of recovered speech recognition performance. The limitations of the online-switching dialog strategy are at a noise level of 80 dB(C), corresponding to a busy urban street with high traffic noise.

[5]Umwelt-Bildungs-Zentrum Steiermark 2008, www.ubz-stmk.at

2.6 Summary

As reasoned in [117] dialogs should be interpreted in terms of informational alignment rather than information transfer. Thus, handling miscommunication is essential for a spoken dialog system. As, according to linguistic models, miscommunication can be divided into three different categories occurring on three different states of understanding, these states are transferred to HRI as states of information retrieval. The states in turn, have been embedded into the second stage of a developed cognitive framework, consisting of three consecutive stages (see Figure 2.2), derived from cognitive theories from social psychology.

Thus, a cognitive framework for proactive information retrieval is developed, and four different dialog strategies are modeled and evaluated in two different experimental settings: Firstly, a fully automated (**FA**) indoor experiment is conducted, where each dialog strategy is evaluated with respect to user experience based on a questionnaire. Secondly, the experiment is replicated within an outdoor Wizard-of-Oz (**WOz**) setting, where three different types of requests for handling miscommunication are additionally employed by the wizard and evaluated in combination with each dialog strategy.

Finally, an online-switching dialog strategy is developed and implemented in a dialog system in order to adapt to varying speech recognition performance by calculating an online-confidence score that initiates switching to a more closed dialog strategy as soon as the confidence decreases below a critical threshold during an interaction. In case the online-confidence increases again, the mechanism switches back to a more open dialog strategy to allow for more natural HRI than by using only closed prompts. An evaluative experiment shows that information extraction from HRI can be kept up in this way instead of aborting an unsuccessful interaction without the switching mechanism. Thus, the benefit of the online-switching dialog strategy is that the robustness against environmental disturbances is improved by increasing the bandwidth of acceptable environmental noise, in which successful information retrieval is possible, from 50 dB(C) up to 70 dB(C) by the application of a switching mechanism. This strategy leads the dialog in a more closed strategy where needed, and switches back to a more open dialog strategy in case of recovered speech recognition performance. The limitations of the proposed online-switching dialog strategy are at a noise level of 80 dB(C), corresponding to a busy urban street with high traffic noise. An outdoor evaluation of the online-switching dialog strategy is pending. Because of the hypothesized marginal traffic noise, a **Closed Dialog** strategy was used in the outdoor field trials of the IURO-project, as described in Chapter 4, Section 4.3.

Since in this chapter, informational alignment was explored in terms of proactive retrieval of missing task-information from natural language HRI, the following chapter is concerned with emotional alignment with regard to proactively triggering prosocial behavior in terms of increased empathy and helpfulness towards a robot.

3 Triggering Prosocial Behavior towards a Robot

In the preceding chapter proactive retrieval of missing task-information from natural language HRI has been explored in terms of informational alignment with the user, resulting in a framework for proactive information retrieval and switchable dialog strategies, capable of handling miscommunication in outdoor environments with varying conditions for speech recognition. In this chapter, emotional alignment is explored with regard to proactively trigger prosocial behavior in terms of increased helpfulness towards a robot, see Figure 3.1.

In applications where (robotic) systems provide human users with information [75, 76, 128], it is self-evident that the user keeps up the interaction as long as all requested information is provided. However, when a robot is asking humans for missing task-knowledge [2, 28, 102, 106] it cannot be taken for granted that humans are interested to interact or even help a robot without any benefit. Thus, this chapter focuses on the development of an emotional adaption approach to proactively trigger increased helpfulness towards the robot in task-related HRI. According to social-psychological predictions of prosocial human behavior, the approach aims at inducing not only empathy, but paired with a feeling of similarity, e.g., in personal attitudes, in human users towards the robot. This is achieved by the development of two differently expressed emotional control variables: by an explicit statement of similarity before task-related interaction, and implicitly expressed by adapting the emotional state of the robot to the mood of the human user, such that the current values of the human mood in the dimensions of pleasure, arousal, and dominance (PAD) are matched. The thereby shifted emotional state of the robot serves as a basis for the generation of task-driven emotional facial and verbal expressions, employed to induce and sustain high empathy towards the robot throughout the interaction. The approach is evaluated in a user study utilizing an expressive robot head. The effectiveness of the approach is confirmed by significant experimental results. An analysis of the individual components of the approach reveals significant effects of explicit emotional adaption on helpfulness, as well as on the HRI-key concepts anthropomorphism and animacy.

The innovation of this chapter consists in the development and evaluation of a novel emotional adaption approach to proactively increase altruistic helpful behavior towards a robot in a persuasive way. In contrast to most state-of-the-art approaches, the goal is an improvement of HRI not only for the benefit of the user, but also for the robot to make use of targeted behavior control deduced from social psychology in order to trigger human helpfulness while fulfilling its task. Contributions are the direct transfer of social-psychological principles from HHI to HRI into a behavior control model for a robot, combining two independently working emotional control variables, implementable in a robotic system. A key challenge is the development of social-psychological "drivers" for empathy towards a robot in a first step, and for the induction of a feeling of similarity to the robot in a second step, which are known in social psychology to trigger altruistic forms of helpful behavior. Another challenge is to develop an objective behavioral measure for altruistic helpfulness towards a robot that is repeatable and thus kept constant over all experimental trials.

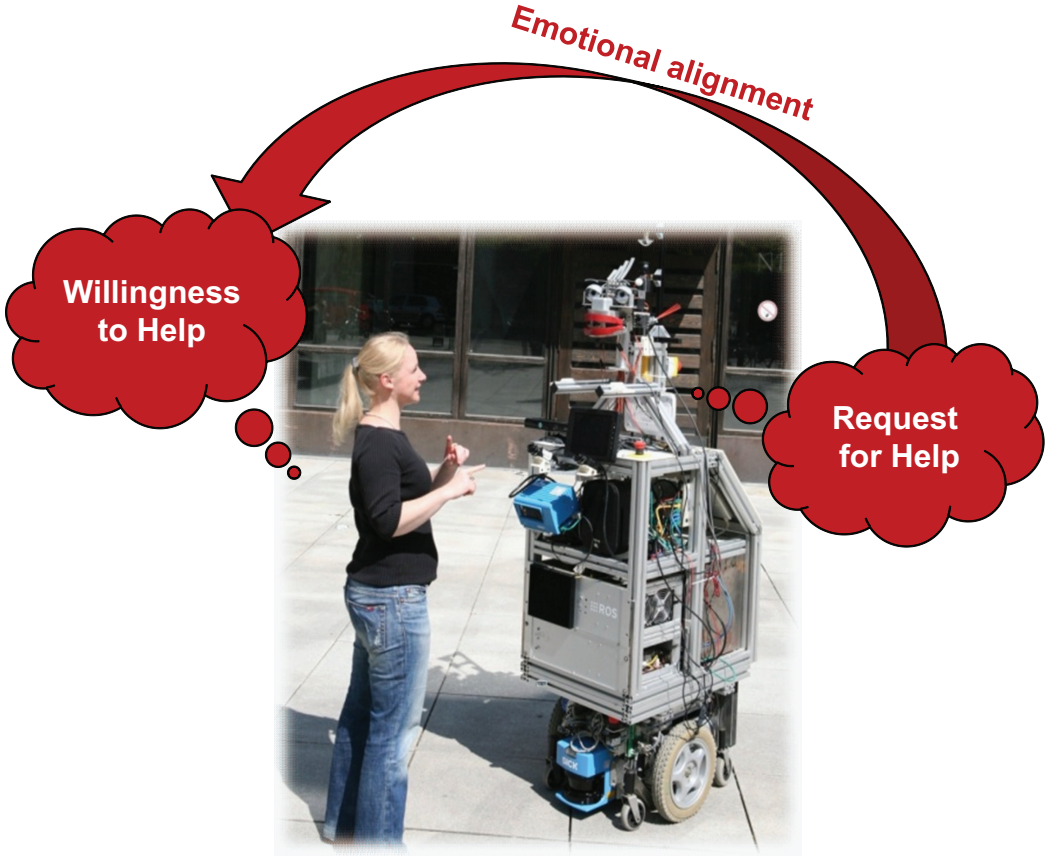

Figure 3.1: Triggering prosocial behavior towards a robot

The remainder of this chapter is organized as follows: In Section 3.1, a problem description is given. Section 3.2 is concerned with the induction of empathy towards the robot by means of a comparative evaluation of facial expressions shown to the user during interaction, animated in different emotional ways. The experimental evaluation reveals not only increased empathy towards the robot for the animation of emotional facial expressions in a socially adaptive way, but also improvements of the perceived subjective task-performance of the robot as well as increased user acceptance. In Section 3.3 an emotional adaption approach is developed to induce not only high empathy, but also a feeling of similarity in the human user in order to trigger altruistic helpfulness towards the robot. The approach is based on theories from social psychology resulting in a behavior control model incorporating two different emotional control variables, explicit and implicit emotional adaption, that are applied and evaluated in combination and as single components in a user study, described in Section 3.4. A behavior measure for altruistic helpful behavior is developed for the experiments. Experimental results show a significant increase in helpful behavior towards the robot for the application of the full approach while significant effects for the emotional control variable of explicit emotional adaption are approved on helpfulness.

3.1 Problem Description & State of the Art

In any interaction, emotions are an important issue. In 1995, Picard introduced the term "Affective Computing" [115]. It describes a form of computing that "relates to, arises from,

or influences emotions". Picard emphasized that this might lead to increased performance and decision making for the computer, stressing the importance of such ideas. Today, a large amount of works incorporate this idea. Two main aspects of affective computing are systems detecting emotions in the human user or conversation partner, and systems showing emotions themselves. The detection of emotions and its use in behavior control is treated in several works, e.g., e-learning systems [3], pedagogical agents [46], driver assistants [4], virtual agents [68], psychological assistance [72], etc. However, the effectiveness of automatic emotion recognition is still very limited and the connection between perceived and real emotions remains an open issue. Also in HRI, emotion recognition, expression, and emotionally enriched communication and closed-loop behavior control have gained strong attention during the last two decades [80, 94, 112, 119, 129]. In human-human interaction (HHI), empathy is crucial for socialization. This ability is already developed in infants [19] and dysfunctions in feeling empathy might lead to social deficits, as observed in autism [40]. In the course of several social psychological studies investigating inter-human empathy, the experimentally induced extent of empathy has successfully been manipulated via similarity of personal attitudes between the subjects [11], [84]. Additionally, studies on unconscious mimicry present findings on the importance of facial mimicry in social interaction. Chartrand and Bargh [34] showed in an experiment that behavioral mimicry ("the chameleon effect") has a significant effect on the interaction and increases empathy towards the interaction partner. There is evidence that feeling empathy for others can be traced back to the mirror neuron system [40], [53], [69], triggering emphatic emotion by deriving the emotional state from facial expressions, and thus involves neural activity in the thalamus and cortical areas responsible for the face. Models from social psychology [51] describe how humans predict events as well as the behavior of other humans [54] and have certain expectations on how a conversation partner will react. The analysis of HRI from a social-psychological perspective does not only reveal important implications for hardware design [171], but can also provide a framework and guidelines for the design of robot communication and behavior [79]. In the research field of "Persuasive Technology" [50], non-robotic technologies, such as internet services or mobile devices, are investigated and developed to change attitudes or behaviors of human users by means of non-coercive persuasion and social influence. One example is an interactive mannequin for shop windows to persuade bypassing customers in order to extend the perceived time they stay in front of a shop window [121].

Most works on social robots are guided by the premise that robots should adapt to humans in order to facilitate intuitive interaction. Nonetheless, proactivity of robots is equally important in order to realize social interaction or to even enable the robot to accomplish its tasks by proactively triggering human behavior [108, 109].

Possible application scenarios are cases where the robot needs the help of humans to achieve a given objective. In the "Interactive Urban Robot (IURO)" project[1], a social robot is developed, capable of proactively acquiring directional information from humans in order to achieve its objective to navigate to goal locations in urban environments, e.g. to perform fetch-and-carry tasks like medicine delivery to its human user. By triggering helpful behavior of humans, IURO is robust against dynamic environmental changes, which cannot be pre-programmed.

[1] see http://www.iuro-project.eu

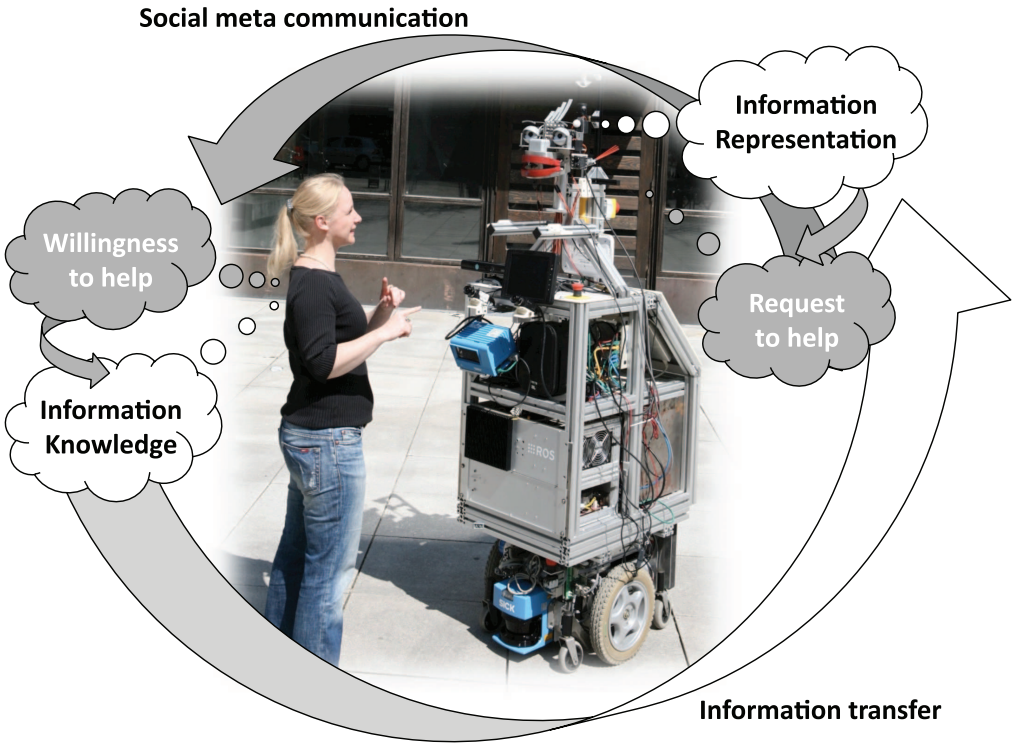

Figure 3.2: Social interaction components as a motivational basis for task-related HRI

Thereby, the request of the robot for help as well as the willingness of the human to help, can be regarded as social meta communication that serves as a motivational basis for information transfer, e.g. missing task knowledge, see Figure 3.2. Thus, for application scenarios where a robot relies on prosocial behavior of humans, triggering human helpfulness is a social sub-task for the robot, necessary to be achieved in order to fulfill its task.

The willingness of passers-by to support robots asking for directions in public spaces has been investigated in previous outdoor-experiments: According to Weiss et al. [165], "the large number of people interacting arises from the fact that many of the interactions were started by curious passers-by". However, in a long-term perspective, service robots might no longer be a novelty in public spaces and curiosity may pass into rejection.

In this context, this chapter describes a behavioral approach and integrated system to trigger more prosocial human reactions in terms of increased helpfulness towards a robot. The approach is developed by transferring social- psychological principles from human-human interaction to HRI. The main idea is to trigger helpfulness in a behavioral way, using both, explicit and implicit communication modalities to create empathy and a feeling of similarity.

A number of studies have already been conducted which employ empathy and similarity as factors in human-robot or human-computer interaction to manipulate the attitude of users towards an artificial agent. In relation to this work, they can be categorized whether the artificial agents are used to express empathy [38, 99, 110, 113, 118, 149, 155] or induce it in the user [113, 114, 124] as proposed here.

Empathic expressions by the agents are mostly utilized to enhance the user experience and thus provide a benefit to the user. Depending on the correct situation awareness and choice of expression, the empathic reactions can be comforting to the user [118], build trust [38], enhance the system perception by the user [99, 113], enhance the subjective task performance [149] and meet user expectations [110]. Thereby, the expression of empathy is either based on empirical data [99], a theoretical model [149] or both [113]. Visual [149], auditory [149, 155] or physiological [118] cues or training data from observations of HHI [99] are used to evaluate the situation of the user and to express an emotion that is similar to the estimated emotional state of the same.

Another approach is to induce empathy in the user by emotional alignment with the same. This is, for example, achieved via facial mimicry [124] or character appearance [114]. While the induction of empathy can enhance the system perception by the user, it is also possible to facilitate altruistic behavior. An example is the work by Paiva et al. [114], in which the character design of experimentally mistreated virtual agents provides similarity to the user and thus the educational aspect of bullying prevention is expected to be raised.

In this chapter, the approach is to proactively trigger altruistic helpful behavior towards a robot in situations, where helpfulness can be avoided by walking away. Unlike other state-of-the-art approaches, the benefit of empathy and similarity is not user-oriented, i.e. not restricted to the internal states of human users in terms of increased user experience and/or educational success. In contrast, the presented approach is task-oriented with regard to directly trigger external human behavior that benefits the robot to better fulfill its task. This is achieved by transferring theories from social psychology [11, 51, 84] to HRI, predicting for situations providing a possibility to avoid helpfulness, that altruistic helpful behavior cannot be achieved via empathy alone, but only paired with a feeling of similarity in personal attitudes and/or characteristics. Hence, the proposed approach focuses not only on the induction of empathy but also on the induction of similarity felt by a human user towards a robot.

As a first step towards this goal, the extent of induced empathy is explored with regard to different ways of animating emotional facial expressions during task-related interaction with a human user, as described in the following section.

3.2 Inducing Empathy towards a Robot

In this section, emotional facial mimicry is applied in a human-robot communication scenario.

The influence of behavioral mimicry has been subject to studies in the field of human-human-, human-agent- and human-robot-interaction. Related work has already shown the transferability of inter-human-findings to virtual agents and social robots. Gratch [63] reports on “virtual rapport” with virtual agents, showing benefits of mirroring head movement and posture shifts resulting in increased speaker engagement and improvements on the interactional level compared to unresponsive agents. The work of Bailenson and Yee [5] on “digital chameleons” concludes that the mimicking head movements of embodied virtual agents are viewed as more persuasive and likeable compared to agents with prerecorded movements. In the field of social robotics, Kanda et al. [74] could improve route guidance interactions with a robot by incorporating cooperative body movements (e.g. synchronization of arm movements), enhancing both reliability and sympathy. Riek et al. [122]

studied the effects of automatic head gesture mimicking with a chimpanzee robot. The listening-behavior of the robot to the subjects is varied in the conditions of either mimicking all head gesture, only nodding, or no mimicking, which result in different levels of interaction satisfaction.

This work extends the state-of-the-art by explicitly evaluating the extent of empathy, induced by facial expressions in contrast to interactive impacts of head, arm or body gestures. The facial expressions are generated automatically and online during HRI. Further, two new questionnaires are developed to evaluate the extent of situational empathy and subjective task-performance.

The core idea is to compare the extent of induced empathic emotions towards a robot during task-related interaction between two different approaches of emotional facial expressions animation, developed by Sosnowski et al. [60]. The first approach is to mirror the facial expressions a user shows in the course of a communicative task in order to trigger the mirror neuron system of the user and thus evoke empathy for the robot. The second approach is to animate the facial expressions shown by the robot according to a "social motivation model" that aligns the interactive smiling reactions between the user and the robot in a socially adaptive way, as described more detailed in Subsection 3.2.1. In the presented user-study, the subjects are asked to rate their situational empathy and the subjective task performance of the robot after playing an interactive question-response game with a robotic head, showing the varying facial expressions according to the conditions of neutral facial expressions, mirroring facial expressions and according to a social motivation model (SMM).

The user-study and its results are also published in [60].

3.2.1 System and Methods

The system, developed by Sosnowski et al. [2], consists of several modules, as can be seen in Figure 3.3.

A module for the recognition of facial expressions, developed by Mayer et al. [3], and a facial expression display module work continuously in parallel, permanently tracking and aligning the facial expressions of the robot with the facial expressions of the user. Further, the robot head turns the neck to focus the face of the user. Text-to-speech is integrated by Blume et al. [4] to communicate the questions of the used question-response game "Akinator" (see: www.akinator.com) to the user. The robot head parses each question to generate adequate lip movements. A speech recognition module passes the verbal responses of the human back to the robot, that sends them back to the Akinator-game via a web API.

The modules are interconnected with a suitable communication backbone based on the Real-time Database (RTDB) introduced by Goebl and Färber [55]. It provides a shared-memory implementation with integrated data storing and is able to handle large amounts of data in real-time, required for instance by the vision-based components of the system.

[2] Dipl.-Ing., Institute of Automatic Control Engineering (LSR), Department of Electrical Engineering and Information Technology, Technische Universität München

[3] Ph.D., Intelligent Autonomous Systems Group, Department of Computer Science, TU München

[4] Dipl.-Inf., Institute for Human-Machine Communication, Department of Electrical Engineering and Information Technology, Technische Universität München

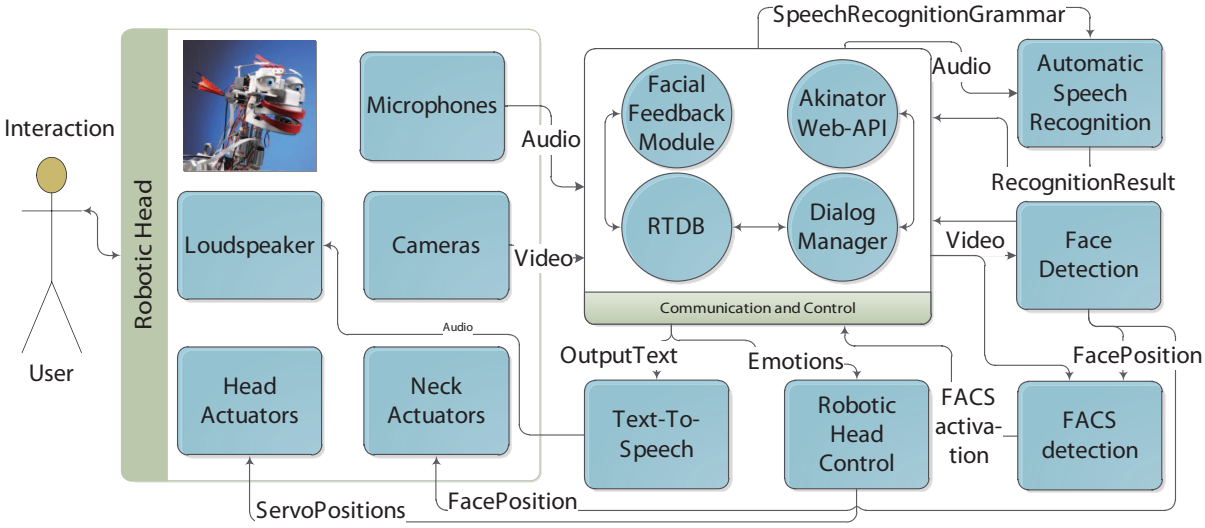

Figure 3.3: Overview of the integrated Modules in the System [60]

The Robotic Head EDDIE

For the experiment the EDDIE (Emotion Display with Dynamic Intuitive Expressions) - head is used [139], an emotionally expressive robot head designed as an interaction partner with 23 degrees of freedom and mixed anthropomorphic and zoomorphic features, see Figure 3.4.

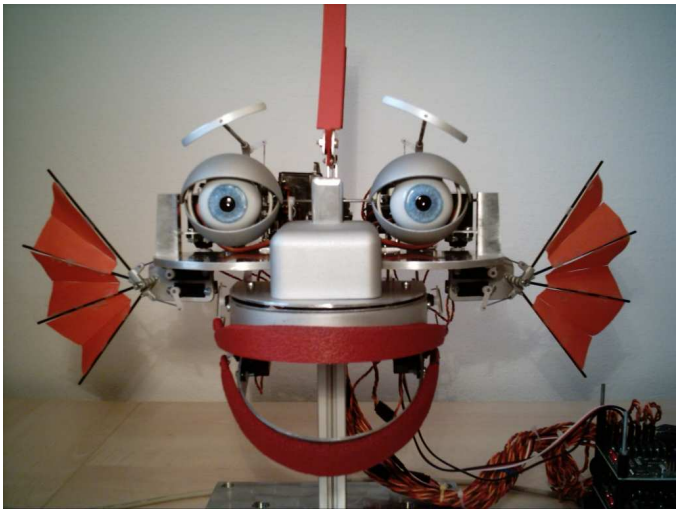

Figure 3.4: The robot head EDDIE [139].

By choosing additional animal-like characteristics, the robot is intended to not provoke disproportionate expectations concerning its social abilities [86]. The basic functionalities of EDDIE are: eye balls 2 DoF, eyelids 2*1 DoF, ears, 2 DoF, mouth/jaw 1 DoF, lips 2*2 DoF.

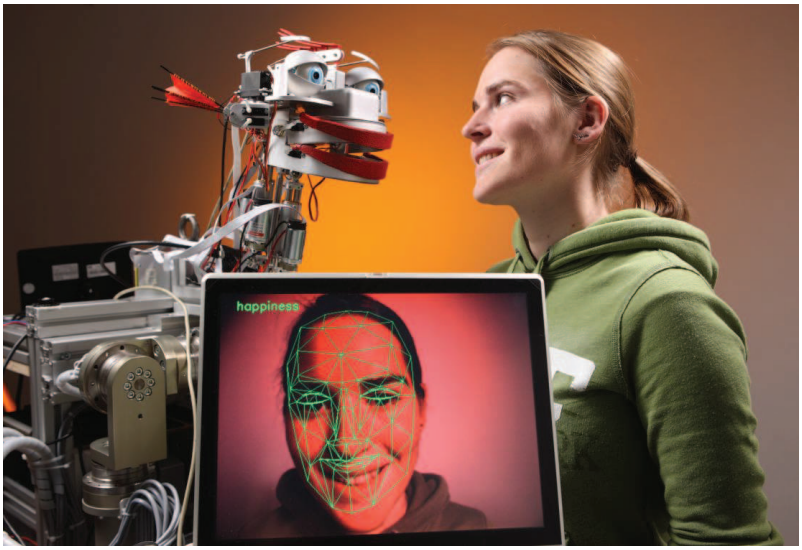

Figure 3.5: The face model is fitted to each image in order to estimate the currently visible facial expression. Photo: Kurt Fuchs

With this head, 13 out of the 21 emFACS [5] action units can be displayed.

EDDIE can be operated on various levels of the control hierarchy, ranging from high-level control of the emotional state to directly sending motor commands. In this case, direct control of the action units is used to achieve mirroring of facial expressions. The visualization of speech is done by parsing the text-to-speech and generating a set of visemes to accompany the speech output.

Facial Expression Analysis

For tracking the face of the human interaction partner and thus showing the focus of attention, the robot head is equipped with a pan-tilt unit by *Directed Perception* as a neck. The facial expressions of the human are analyzed from camera images, see Figure 3.5

According to Mayer et al. [60], the facial components are considered separately to determine intra-face movements like raised eyebrows or an opened mouth. The system calculates activation intensities of several FACS action units during the interaction. Furthermore, the position of the human face is determined in 3D space, enabling the robot head to focus on the user by turning the head at the neck.

A model-based technique is used to determine the exact location of facial components such as eyes or eye brows in the image. The Candide-3 face model is a wire-frame model consisting of 116 anatomical landmarks [1]. Its parameter vector describes the face pose in 3D space and the face shape. In order to extract action unit activations for a single image, model parameters that match the image content are calculated. For instance, if the user visible in an image is smiling, the model parameters should reflect raised lip corners. The approach requires a neutral reference image of the user to calculate corresponding model parameters. No prior knowledge of the image content or the user is available. In subsequent images, the model is tracked and model parameters are compared with the neutral face to determine action unit activations.

[5] emFACS is a subset of the facial action coding system, including only action units which are involved in emotional facial expressions

The action units recognized by the analysis components and synthesized by the robot are AU2 (outer brow raiser), AU4 (brow lowerer), AU5 (upper lid raiser), AU7 (lid tightener), AU13 (lip corner depressor), AU26 (yaw drop), AU42 (eyes closed). Based on this information, the robotic face calculates a corresponding facial expression for displaying an appropriate reaction.

Social Motivation Model

The implementation of the social motivation model is based on a reduced version of the Zurich Model of Social Motivation [64] and can describe the effect of smiling and other facial expressions based on the motivational and emotional state of a human or agent [20]. In a concise description, the model combines three motivational subsystems regulating security, arousal, and autonomy. These systems are homeostatic. The autonomy regulation has a special role, since it is coupled to security and arousal. One of the main assumptions in this model is that smile reactions are the result of a decline in autonomy, meaning that smiles are a reaction to external disturbances of the homeostasis, like social distance changes, environmental changes or conflicts, etc. Changes in the respective subsystems lead to characteristic facial expressions, which in superposition result in the overall facial expression. With this model, an agent is able to react to various, even unknown, situations as long as the parameters for security, arousal and autonomy can be extracted. For more detailed information on the composition of the social model, please refer to [20].

For this experiment, the model is extended by Sosnowski et al. [60] to use the facial expressions of the human interaction partner as an input. Smiling at the robot increases the security state, thus resulting in a smile reaction of the robot. Detected arousal increases the level of arousal in the system and angry or very stern looks can be interpreted as a challenge to the autonomy. All these inputs provoke a reaction of the robot that is quite similar to the input signal, but the reaction is delayed by about one second, due to the frame-rate of the facial analysis and model-internal time constants, and influenced by the actual motivational state of the robot.

Dialog and Akinator

In order to provide structure and context to the ongoing dialog, a speech-interface to the "Akinator" (see www.akinator.com), a web-based application that is usually executed in a browser, is integrated by Blume et al. [60]. In this dialog, the user is asked to think of a person. Then, the computer tries to guess this person by asking several questions. The person may be a real or fictional person, currently living or historical, taken from literature, the media or public live. To answer Akinator's questions, a set of fixed answers is presented by the system. The set of answers is the same for every question and consists of: "Yes", "Probably" / "Partially", "I don't know", "Probably not"/ "Not really", and "No". Example questions asked by the Akinator are: "Is your character a girl?", "Does your character live in America" or "Does your character really exist?".

In order to create a dialog with the robot head, text-to-speech is used to present Akinator's questions acoustically to the subject and speech recognition is utilized to retrieve the answers.

A dialog manager keeps track of the ongoing communication to estimate when a response of the human user or the machine is expected by the dialog partners. The complete dialog structure is implemented in a first-order logic representation. Tasks to be solved are represented by predicates with variables. These variables represent information to

be determined during the dialog. Equivalence rules on these predicates are specified to navigate through the dialog by splitting a task into several subtasks. Evaluating predicate truth values and binding variables models real-world interaction.

3.2.2 Experiment IV: Evaluation of Emotional Impacts of Facial Expressions-Animation on task-related HRI

In order to evaluate the impact of different ways of emotional facial expressions-animation, an experimental setup is created for the subjects engage in a dialog with the robot head EDDIE. The web-based gaming application "Akinator" serves as a backbone for the dialog structure. In this game, the robot tries to guess a person thought of and chosen by the subject by asking various questions about the person. During this task-related interaction the robot reacts in various ways to the facial expressions of the human, either ignoring them, mirroring them, or displaying a facial expression based on the psychological model for social awareness, as described in Section 3.2.1. In which way this robot behavior influences the human perception of the interaction is investigated by questionnaires. The hypothesis is that the robot behavior during interaction heavily influences the extent of empathy by a human towards a robot, as well as perceived subjective task-performance, with the adaptive modes leading compared to the non-adaptive mode.

A key assumption for the experiment on the impacts of how emotional robotic facial expressions are animated, as described in this section,is that the single facial expressions of the robot are interpreted correctly by the human subject. This assumption is strengthened by the findings of a pre-evaluation, as described in detail in Mayer et al. [97] and Sosnowski et al. [140].

During the interaction, EDDIE speaks and tracks the person while acting according to one of three possible conditions. Thus, the experimental subjects are divided into different groups depending on the following experimental conditions:

1) Neutral: EDDIE displays no facial expressions

2) Mirror: Eddie displays the subject's facial expressions

3) Social motivation model (SMM): EDDIE displays facial expressions according to its internal system-theoretic model of socially-adaptive smiling

After the interaction each subject fills in a computer-randomized questionnaire as described in Section 3.2.3.

The main goal of the study is to reveal if mirroring and/or socially adaptive facial expressions in conditions 2 and 3 induce empathy towards a robot and if the user grades the subjective performance of the robot accordingly higher than in condition 1. Further, the study aims to unveil possible impacts on HRI regarding the five key concepts anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety [9], as well as possible influences on user acceptance [66]. Accordingly, the assumed interrelations of user acceptance and HRI key-concepts with and between empathy and subjective system-performance are investigated.

3.2.3 Experimental Design & Measures

For the experimental setup a quiet room with controlled lighting conditions was chosen. The robotic head was placed on a table to be at approximately eye-level with the subjects that were seated in front of the robot, with a microphone placed in front of them on the table to ensure a low error rate in speech recognition as can be seen in Fig. 3.6.

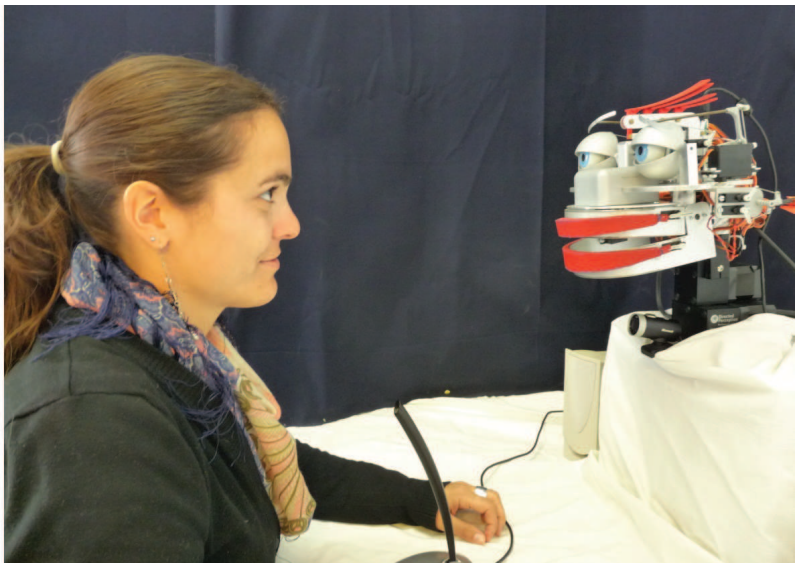

Figure 3.6: Experimental HRI setup [60].

Since the task rating and enjoyment of the interaction would depend on the ability of the robot to correctly understand the answers, the external microphone was preferred over the internal, that would have added to the illusion of speaking to the robot directly. The instructor greeted the person and gave a short introduction on the task and how to interact with the robot. In order to start the experiment, the instructor asked the participant to think of a person of his/her own choice and give a start signal, when done. From this point, the robot started the akinator game, speaking the questions provided by the Akinator API and listening for the answers. A sample round of Akinator can be seen in Table 3.1.

After the game was finished by either the robot guessing the correct person or giving up after too many trials (dependent on the Akinator API, having a threshold influenced by the confidence and the number of trials), the subjects were asked to fill in a computer based questionnaire.

Experimental Measures

The computer-randomized questionnaire consists of two different parts which can be analyzed independently.

The first part consists of five selected constructs based on a "limited model for studies on social abilities or social presence" out of a toolkit for measuring user acceptance of social robots [66]. These constructs are adapted to the requirements of the experimental setting and kept constant with regard to a consistent number of items, i.e. four questions for each construct. Additionally, these five constructs are enhanced by two newly developed constructs, which are proposed to measure the induced scope of situational empathy towards a robot, and the subjective system-performance perceived by the user. These additional

Question	Answer	
	given	expected
Is your character a male?	No	No
Is your character a singer?	No	No
Does your character really exist?	No	No
Does your character fight?	Not really	No
Is your character from an anime?	No	No
Does your character live in America?	No	No
Is your character a human being?	No	No
Is your character an animal?	No	No
Does your character have hair?	No	No
Is your character visible?	Yes	Yes
Is your character a robot?	Yes	Yes
Has your character played in 'Star Wars'?	Yes	Yes
Is your character yellow?	No	No
I guess you were thinking of: R2D2		

Table 3.1: Sample dialogue of a game of akinator, looking for R2D2

constructs are to reveal supposed interrelations to the other constructs on user acceptance and thus enhance this existing toolkit.

The second part of the applied questionnaire consists of the "godspeed" questionnaires [9] to evaluate the "five key concepts of HRI": anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety.

Hence, the questionnaire evaluates the interaction on three different dimensions: 1) Empathy and subjective performance as proposed extension of 2) user acceptance, and 3) the key concepts of the godspeed questionnaires.

In the following the two parts of the questionnaire are described in detail.

Empathy and Subjective Performance

For measuring both constructs, the scope of induced situational empathy on the one hand, and subjective system-performance on the other hand, this part of the first questionnaire is divided into two different paths depending on the objective system performance in the task, i.e. if EDDIE is successful (a) or not (b) in guessing the thought-of person. Thus, subjective performance can be compared to objective performance in order to draw conclusions on possible interrelations due to the scope of induced empathy. Therefore, the subjects are asked to respond to different statements including positive, negative and inverted formulations for sharing happiness or sadness with EDDIE corresponding to the task-success or -failure of EDDIE as shown in Table 3.2.

Users can reply to these statements on a five-item Likert scale (1=strongly disagree to 5=strongly agree). For analyzing the answers correctly, the scale for negatively formulated items, e.g. questions 4a) and 4b), has to be inverted afterwards.

User Acceptance

Heerink et al. [66] extended the Unified Theory of Acceptance and Use of Technology (UTAUT) model [156] by several constructs in order to adapt this model to the specific

requirements of evaluating social robots. Given experimentally validated interrelations between several constructs, the five selected constructs include:

Trust: The belief that the system performs with personal integrity and reliability

Perceived Sociability: The perceived ability of the system to perform sociable behavior

Social Presence: The sensing of a social entity when interacting with the system

Perceived Enjoyment: Feelings of joy or pleasure associated with the use of the system

Intention to Use: The outspoken intention to use the system over a longer time period

The questionnaire evaluates each construct by four different statements, as presented in Table 3.2. In order to reduce acquiescence bias some items are negated and thus invert the scale.

Again, the subjects rate the randomized statements on five-item Likert scales (1=strongly disagree to 5=strongly agree). As the statements for user acceptance and their constructs are independent from the system performance this questionnaire is not divided into different paths if EDDIE was successful (a) or not (b) in guessing the person thought of. Nevertheless, it is analyzed if interrelations to subjective task performance exist.

Godspeed Key Concepts [6]

"A series of questionnaires to measure the user's perception of robots" combines five consistent and validated questionnaires based on 5-point semantic differential scales as a standardized metric for the "five key concepts in HRI" [9]:

Anthromorphism: how natural the robot appeared

Animacy: the liveliness of the robot

Likeability: how pleasant the robot appeared

Perceived Intelligence: how the mental abilities of the robot were perceived

As recommended, the items are randomized so as to hide the different concepts and hence mask the intention. However, in order to avoid capturing changes of the emotional state of the subjects while filling in the questionnaire, in this study the emotional state of the user is measured directly after the interaction with EDDIE, and thus the three questions of *Perceived Safety* constantly set up the beginning of the overall questionnaire.

3.2.4 Experimental Results

Results are deduced from the experimental evaluation including 55 subjects (40 male and 15 female, between 21 to 60 years with an average age of 28.8).
The distribution of the subjects over experimental conditions was 13 for **Neutral**, 17 for **SMM**, and 25 for **Mirror**.

[6] Open source version, see http://www.bartneck.de/2008/03/11/the-godspeed-questionnaire-series

	Situational Empathy
1a)	I am happy that EDDIE guessed my person.
1b)	It's a shame EDDIE didn't guess my person.
2a)	I would have been proud if Eddie hadn't guessed my person. (inverted)
2b)	I'm proud EDDIE didn't guess my person.
3a)	It would have been a pity if EDDIE didn't guess my person.
3b)	It would have been nice if EDDIE had guessed my person.
4a and b)	I would feel sorry for EDDIE if someone tried to destroy it at that moment, thus I would try to prevent it.
	Subjective Performance
1a)	I was impressed by how fast EDDIE has guessed my person.
1b)	I had the feeling that EDDIE nearly guessed my person.
2)	EDDIE has shown a good performance.
3)	I think that EDDIE has worked efficiently.
4a)	It took EDDIE long to guess my person. (negated)
4b)	It took EDDIE too long to guess my person. (negated)
	Trust
1)	I would believe EDDIE if he gave me advice.
2)	EDDIE is inspiring confidence.
3)	I feel that I can trust EDDIE.
4)	I do not trust EDDIE's statements. (negated)
	Perceived Sociability
1)	I like EDDIE.
2)	EDDIE's mimic and verbal statements fit together well.
3)	EDDIE was a good conversation partner.
4)	EDDIE's behavior was inappropriate. (negated)
	Social Presence
1)	I had the feeling that EDDIE really looked at me.
2)	I could imagine EDDIE as a living being.
3)	Sometimes it felt like EDDIE had real feelings.
4)	EDDIE's behavior was not humanlike. (negated)
	Perceived Enjoyment
1)	It was fun to interact with EDDIE.
2)	The conversation with EDDIE was fascinating.
3)	I consider EDDIE to be entertaining.
4)	It's boring when EDDIE interacts with me.(negated)
	Intention to Use
1)	I would like to interact with EDDIE more often.
2)	I would take EDDIE home with me.
3)	I would like to play again with EDDIE within the next few days.
4)	I could imagine interacting with EDDIE over an extended period of time.

Table 3.2: Questionnaires for User Acceptance on a 5-point Likert scale, extended by two constructs on Empathy and Subjective Performance

Table 3.3: User Acceptance: Mean ratings (rated on Likert scales from 1 = strongly disagree to 5 = strongly agree) with standard deviations (in brackets) of each construct and total scores within conditions

Construct	Condition		
	Neutral	Mirror	SMM
Situational Empathy	3.1(1.3)	3.7(1.1)	4.4(0.8)
Subjective Performance	2.8(1.2)	3.4(1.0)	4.1(0.9)
Trust	3.0(0.6)	3.3(0.8)	3.7(0.5)
Perceived Sociability	3.2(1.0)	3.6(1.0)	3.9(0.7)
Social Presence	2.8(0.6)	2.8(0.7)	2.9(0.7)
Perceived Enjoyment	2.8(1.4)	3.9(1.2)	4.2(0.7)
Intention to Use	3.0(1.3)	3.5(1.0)	3.9(1.0)
Total Score	2.9(1.1)	3.5(1.0)	3.9(0.8)

Regarding reliability, coefficients of internal consistency are calculated with Cronbach's α for the items of the novel constructs on *Empathy* and *Subjective Performance*. As a solid construct should create an Cronbach's $\alpha > .70$ all items of both novel constructs showed good reliability with Cronbach's $\alpha = .82$ for *Empathy*, and Cronbach's $\alpha > .86$ for *Subjective Performance*. Since the selected constructs for user acceptance and of the Godspeed questionnaires are previously evaluated [9, 66] reliability and internal consistency are assumed.

Significance level for all performed tests was set to $\alpha = .05$. According to the results of Kolmogorov-Smirnov tests, normal distribution could be accepted for the total scores of all constructs, except *Perceived Enjoyment*. Thus, this construct has to be analyzed non-parametrically. Parametric comparisons and correlations are performed for all other constructs.

An analysis of variance (ANOVA) revealed significant differences between the conditions for *Empathy* ($F = 5.35, p = .008$), *Subjective Performance* ($F = 6.48, p = .003$), *Trust* ($F = 4.47, p = .016$), and *Likeability* ($F = 3.73, p = .031$). Thus, a post-hoc analysis could be conducted between the three conditions. Accordingly, the assumed significance level was divided by three and thus adjusted to $\alpha = .016$. Paired t-tests revealed significant differences between **Neutral**- and **SMM** conditions for *Empathy* ($t = -3.01, p = .007$), *Subjective Performance* ($t = -3.51, p = .002$), and *Trust* ($t = -3.30, p = .003$). Between **Neutral**- and **Mirror** condition one significant difference was found for the godspeed construct *Likeability* ($t = -2.03, p = .062$), and no significant differences were found between the conditions of **SMM** and **Mirror** due to the α-value adjustment. Means, total scores and standard deviations of the five constructs on user acceptance by Heerink [66], and the two additionally introduced constructs on *Empathy* and *Subjective Performance* are displayed in Table 3.3.

Mean values and total scores for the five key concepts in HRI, as derived from the godspeed questionnaires, are depicted in Table 3.4.

Table 3.4: Key Concepts (Godspeed): Mean ratings (rated on Likert scales from 1 = strongly disagree to 5 = strongly agree) with standard deviations (in brackets) of each construct and total scores within conditions

Construct	Condition		
	Neutral	Mirror	SMM
Perceived Safety	3.9(0.8)	3.6(0.6)	3.7(0.5)
Anthropomorphism	2.6(0.6)	2.8(0.5)	2.8(0.7)
Animacy	3.1(0.7)	3.3(0.4)	3.3(0.7)
Likeability	3.5(1.1)	4.1(0.5)	4.1(0.7)
Perceived Intelligence	3.5(0.8)	3.8(0.5)	3.9(0.5)
Total Score	1.1(0.7)	3.5(0.5)	3.6(0.6)

Correlation analysis focused on the five selected constructs on user acceptance, along with the added constructs on *Empathy* and *Subjective Performance*. Correlation coefficients led to the finding that all constructs show significant correlations to each other ($p < .001$), except for *Social Presence* which only correlates significantly to *Trust* ($r = .36, p = .007$).

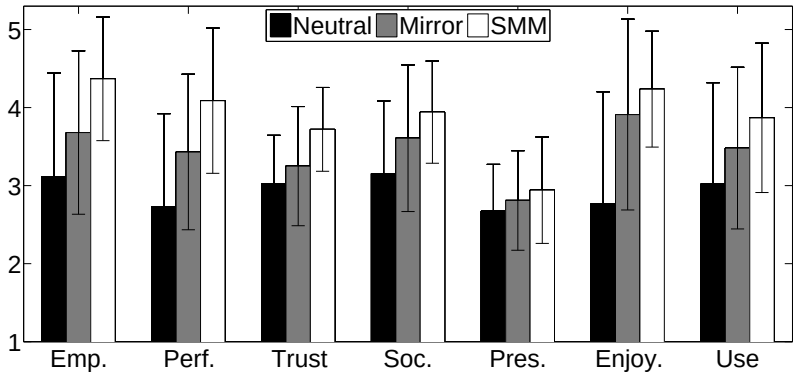

Figure 3.7: Mean values of Heerink's 5 and the introduced 2 additional constructs for 3 conditions: neutral, mirror, and SMM on a 5-item Likert scale from 1 (strongly disagree) to 5 (strongly agree).

3.2.5 Discussion

The experimental evaluation of emotional facial expressions in terms of their effects on HRI provides new insights regarding the possibilities and limitations of their animation. Three different experimental conditions of facial mimicry are implemented in a robotic system and evaluated in terms of user acceptance and key concepts of HRI. Additionally, two new measures for situationally induced empathy and subjective system-performance are introduced and successfully evaluated with regard to their internal reliability and existing

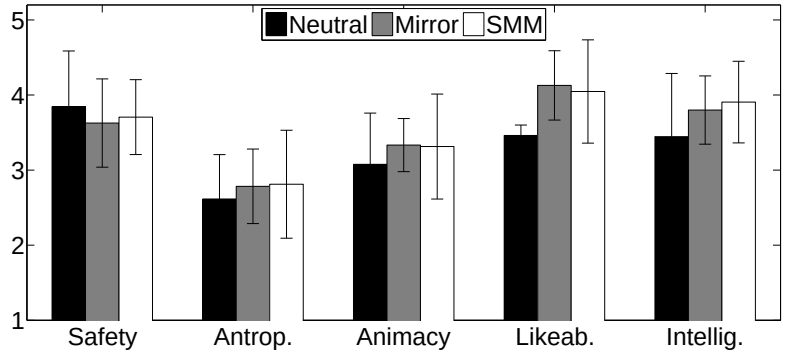

Figure 3.8: Mean values of the 5 godspeed constructs for 3 conditions: neutral, mirror, and SMM on a 5-item Likert scale from 1 (strongly disagree) to 5 (strongly agree).

correlations with user acceptance. In general, results support the initial hypothesis by showing a trend towards a better rating of the mirroring condition compared to the neutral condition, with the social motivation model (SMM) being rated significantly better in most instances. This underlines the importance of social factors to be considered for further refinement of how mimicking should be performed.

The reliability of the newly developed measures for empathy and subjective performance is confirmed and correlations of those with all other constructs on user acceptance, but social presence, are revealed. Also the significance of empathy, subjective performance, and likeability provides evidence for the impact of socially-adaptive animated facial expressions on the interaction. Since EDDIE has a very machine-like appearance it is possible that this may have dominating effects on the construct of social presence: Mean values show no noticeable increase within this construct, and no correlations could be found besides with the construct of trust. However, results indicate that social presence, that is very much bound to being humanlike, is not crucial in order to induce empathy. It is notable that according to Bailenson and Yee [5], the effect persists even when the person being mimicked is fully aware that the mimicker is an artificial agent.

In the following section, the gained insights on induced empathy are re-evaluated by extending this approach to an integrated behavior control model to trigger prosocial behavior in terms of increased helpfulness towards a robot.

3.3 Inducing Empathy & Similarity towards a Robot

In order to trigger altruistic helpfulness towards another human, high situationally induced empathy must be paired with a feeling of similarity towards a person in need of help [51]. Thus, in order to transfer this finding to HRI, similarity is induced by two different ways of emotional expression in the proposed approach: by an explicit statement of similarity before task-related interaction, and implicitly expressed by adapting the emotional state of the robot to the mood of the human user, such that the current values of the human mood in the dimensions of pleasure, arousal, and dominance (PAD) are matched. The thereby shifted emotional state of the robot serves as a basis for the generation of task-driven emotional facial- and verbal expressions, employed to induce and sustain high empathy towards the robot throughout the interaction. In the experimental evaluation

of the approach, these task-driven emotional expressions are kept as a constant over all experimental conditions to sustain high empathy, while the factors of explicit and implicit emotional adaption are varied in a 2x2 between-subjects design in order to reveal their effects on helpfulness, shown by the user towards the robot in task-related interaction, as well as on user experience. In a first step, the user-mood is determined by an initial self-assessment by the human participant to be extended by automatic emotion recognition modules in a later stage. The interaction task is exemplarily designed as a person guessing task. The effectiveness of the approach is confirmed by significant experimental results, deduced from 55 test subjects in previous work (see Section 3.2, and 84 subjects in the presented study. An analysis of the individual components of the approach reveals significant effects of explicit emotional adaption on helpfulness, as well as on the HRI-key concepts anthropomorphism and animacy. The development of the emotional adaption approach and the experimental evaluation of the full approach as well as of the single components as stand-alone emotional control variables are also published in [56, 59, 85].

3.3.1 Background from Social Psychology

In human-human interaction, "prosocial behavior" in terms of altruistically motivated helpfulness and its determinants is a well-studied field of research [18]. The presented approach is inspired by social-psychological studies [11, 84], where a feeling of being "similar" in terms of having something in common with a person in need of help, e.g. in personal attitudes or characteristics, turned out to be a motivational activator for increased helpfulness towards this person, paired with high empathy. Empathy can be defined as "The capacity to know emotionally what another is experiencing from within the frame of reference of that other person, the capacity to sample the feelings of another or to put one's self in another's shoes" [16]. In other words, the extend of personal distress felt by a potential helper when observing a person in need of help depends on the degree of situationally developed empathy for this person, and similarity is the activating factor for either reacting with altruistically or egoistically motivated behavior:

In situations providing a possibility to avoid helpfulness, e.g. by walking away, referred to as "easy means of escape", the feeling of having something in common with the person in need of help (similarity), paired with correspondingly high empathy, activates altruistically motivated helpfulness. Accordingly, the perceived reward for helping is much higher than the reward for walking away, resulting in high helpfulness, see Table 3.5. In contrast, in the absence of similarity, people would only be highly helpful if there was no or only difficult means of escape. This kind of helpfulness is egoistically motivated to reduce one's own discomfort arising from the empathic reaction on the situation.

Thus, in situations with easy means of escape (as given in most HRI-scenarios), people without a feeling of similarity tend to leave the scene showing low helpfulness towards the person in need of help, since this is an equally efficient way of reducing the negative empathic stimulus. The degree of empathy would not play a role in this case [18]. In Table 3.5, the social-psychological predictions on helpfulness are summarized for situations with easy means of escape, considering the influence of similarity, paired with high empathy.

Since in most HRI-scenarios easy means of escape are provided, the approach is to raise the motivation of human users to help the robot, e.g. in public places. According to the findings of social psychology, the approach is to design the interaction in a way to induce similarity between the robot and the user, paired with high empathy towards the same.

Table 3.5: Predictions on helpfulness for situations with easy means of escape according to social-psychological theories [11, 18, 84]

	Low empathy	**High empathy**
With similarity	Low helpfulness	**High helpfulness**
Without similarity	Low helpfulness	Low helpfulness

Hence, in order to increase helpfulness towards a robot, the presented experiments focus on a constant induction of high empathy, paired with the experimentally varied induction of similarity. Constantly high empathy is achieved by emotionally adaptive facial expressions of the robot, as investigated in previous work, see Section 3.2, incorporated in the developed approach. Regarding the induction of similarity, an evaluative variation of two different persuasive emotional control variables, developed earlier as components of the emotional adaption approach [56, 59], is applied. The experimentally evaluated parts of social-psychological predictions and corresponding human target behaviors are marked in gray color in Table 3.5.

For the development of persuasive emotional control variables, all available robotic output modalities should be used. The following subsection provides an overview on explicit and implicit communication modalities with regard to their linguistic background and applications in HRI.

Explicit versus Implicit Communication

In linguistic pragmatics, a distinction is made between explicitly communicated content which is directly said or written, and "implicatures" [30], that enrich and manipulate the pragmatic interpretation of explicitly communicated content. Accordingly, communication modalities are not limited to explicit communication channels like direct verbal or written utterances, but also "silent messages" [100] as implicit communication channels of emotions and attitudes. According to Mehrabian [100] this includes "all facets of nonverbal communication, including body positions and movements, facial expressions, voice quality and intonation during speech, volume and speed of speech, subtle variations in wording of sentences that reveal hidden meanings in what is said, as well as combinations of messages from different sources, e.g., face, tone of voice, words." This holds equally true for HRI, where beliefs about the other's mind are also resulting from interpretation of the other's behavior, that becomes a "sign" of their own minds, by means of implicit and explicit ways of communication [32].

The importance of such "mutual beliefs" in natural language communication is instantiated in the phenomenon of "grounding" [36], meaning that the interpretation of communicated contents has to be at least "approximately correct" in order to achieve successful communication acts, based on a common underlying field of knowledge and/or required actions [152]. Also for artificial social agents, Castelfranchi stresses the importance of a "basic ontology of social action" with special focus on prosocial forms in the mental representations as beliefs and goals of the agent in a social interaction [33].

In the presented approach, focus is set on the adaption of emotional facial and verbal expressions in an implicit and explicit way: An explicit statement of similarity is given by the robot by verbally expressing that it is in the same mood as the user prior to task-related HRI. Implicit emotional adaption is conducted by shifting the base-values of emotion facial end verbal expressions (prosody in speech) towards the user-mood during task-related HRI.

The implicit modality of facial expressions has already been explored in terms of inducing high empathy in previous work and is shortly outlined in the following.

3.3.2 Transfer to HRI: The Emotional Adaption Approach

The basic idea is to induce both, high empathy and a feeling of similarity in a human user towards a robot by adapting to the mood of the user and thus providing the human with the impression of sharing the same emotional state as a starting position for the interaction. To achieve this, the emotional adaption approach is divided into two components which express the adaption to the mood of the user in two different ways: explicitly and implicitly. Explicit expression of similarity is given by stating "me too" when the user was asked about her mood, as outlined more detailed in the following, and in Section 3.3.4. Implicit expression of similarity is generated using facial and verbal emotion expressions during the HRI task execution that are biased using the mood of the human as measured before the interaction. In the implicit case, as described more detailed in the following, and in Section 3.3.5, similarity consists of an initial bias of the emotional state of the robot, based on the user mood. In the course of task-related interaction, this bias serves as a shifted baseline for the generation of task-driven emotional expressions of the robot that are included to induce and sustain high empathy in the human user in accordance with the experimental findings of previous work, see Section 3.2.

As an example for implicit emotional adaption, previous work showed that empathy and other dimensions of HRI could be improved by the emotional animation of facial expressions to the human user [60]. However, a socially adaptive way of reacting to facial expressions, shown by a user during interaction, requires robust recognition and analysis of the facial action units involved, based on camera images [97]. Since the recognition quality may often be impaired by dynamically changing environmental impacts like varying light conditions or unpredictable background-movements which may distract the focus of a face tracker, the approach of emotional adaption additionally includes an explicit emotional adaption method. Hence, the approach is not restricted to implicitly expressed mimicry or prosodic variations in speech, but also applies explicitly uttered statements to induce similarity, modeled according to underlying social psychological principles. Another advantage is the increased robustness against environmental impacts: If bad performance of automatic speech recognition impairs the explicit part of emotional adaption, the approach may still be robust in terms of implicit emotional adaption. Hence, two different emotional control variables for prosocial HRI are proposed, capable of compensating each other with regard to varying recognition performance of speech or facial expressions, as depicted in the developed behavior control model, see Figure 3.9: For the robot, the emotional control cycle starts with the input of the user-mood as starting point for emotional adaption mechanisms. This can be achieved by emotion recognition modules or, as applied in the presented study, by an initial self-assessment of the user. Subsequently, the robot initiates the dialog with the user and applies explicit and/or implicit emotional adaption during the interaction. Thereby, the robot persuades the user to show prosocial behavior, e.g. in terms of increased helpfulness towards the robot.

In the following, the two components of the approach, namely explicit and implicit emotional adaption are explained, and related control variables, as used in the presented experiments, are defined.

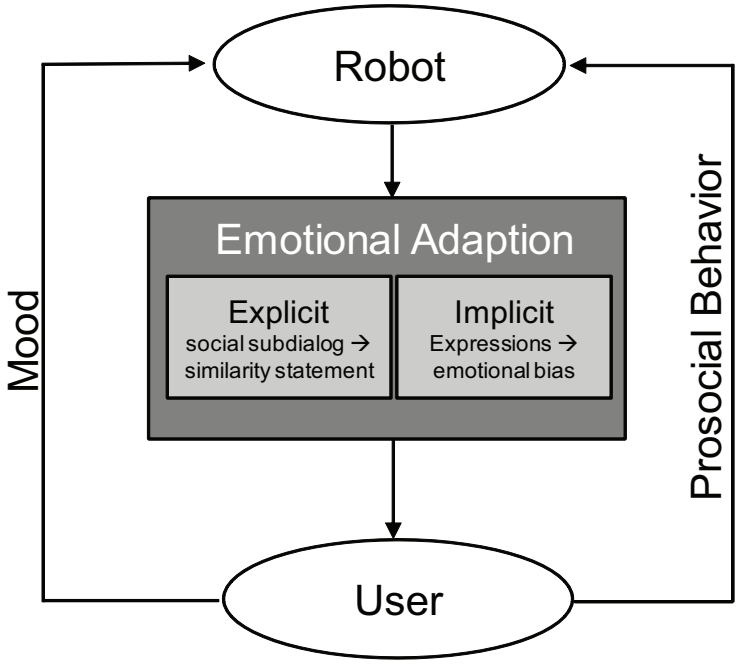

Figure 3.9: Emotional control cycle for prosocial behavior in task-related HRI: After the input of the user-mood the robot persuades the user by explicit and/or implicit emotional adaption to trigger more prosocial behavior in turn.

Explicit Emotional Adaption

Independent of the interactive goal which is expressed later during task-related human-robot dialog, the idea is to implement some small talk to open the dialog and thereby monitor the current mood or other personal attitudes of the user. Thus, an explicitly expressible basis is provided to induce a feeling of similarity between the user and the robot. Thereby, it has to be considered that this may not match the actual mood but only the mood, the user is willing to communicate because of social conventions and rituals during small talk [150]. However, even when communicating with embodied artificial agents, humans build rapport and trust by means of small talk [17]. The instrumentalized form of small talk used in the presented approach is referred to as "social subdialog" in the following, since triggering helpfulness by means of similarity is regarded to be a social sub-task in cases where helpfulness is necessary to fulfill the overall task. In the course of this social subdialog, explicit emotional adaption, and thereby similarity, is created by directly stating a mutuality in an attitude or, as applied in the presented study, in the current mood. Thereby, an impression of having something in common with the user is created.

Accordingly, the emotional control variable of explicit emotional adaption is a directly uttered similarity statement during a social subdialog.

Implicit Emotional Adaption

Existing HRI-applications using implicit communication channels are based on a communicative mechanism in human-human interaction, called "alignment" [116], that leads to adaptive processes between interlocutors which are essential for human-human interactions [49, 83]. One example is an alignment-approach of emotional facial expressions, where a distinction of automatic, schematic and conceptual levels for emotionally adaptive

reactions is made, as partly implemented in the robotic head "Flobi" [39]. In contrast to state-of-the art approaches, this work additionally aims to create a feeling of similarity in users by adapting to their current mood. Thus, an underlying representation of emotional states is needed for both, the generation of facial and verbal expressions, as well as for decoding and adapting to the mood of a user: the Pleasure-Arousal-Dominance (PAD) model [101], where emotions are presented in a continuous three-dimensional space:

- *Pleasure* describes a person's evaluation of the situation, or, put more generally, how content the person is. High pleasure indicates happiness or gratification, while anger and boredom result in low pleasure values.
- *Arousal* states how agitated the social actor is - regardless of whether this a positive or a negative excitation. High arousal values can be found in angry expressions as well as surprised expressions, while low values can, for example, describe a bored expression.
- *Dominance* is defined as "a feeling of control and influence over one's surroundings and others" versus submissiveness, in the sense of "feeling controlled or influenced by situations and others." [101]

Advantages of using PAD are for e.g. the supportive evidence for the three dimensional categorization of emotions [101], the ability to express a variety of emotional states in varying intensities (even subtle forms) and the availability of assessment tools like the semantic differential, described in Section 3.3.5.

For implicit emotional adaption, the approach is to use the human-like modalities of facial and verbal expressions in terms of mimicry and prosody in speech, but can be extended to any emotional non-human-like modalities by related PAD-values. Before implicitly adapting to the mood of the user, the emotional state of the user has to be determined and mapped to the continuous PAD space. Ideally, this can be achieved by emotion recognition modules [98, 161], but at least according to an explicit statement in the course of the social subdialog introduced above, and/or in combination with an initial self-assessment of the user on the PAD dimensions. When this is achieved, the robot shifts its base-PAD values for emotional expressions towards the mood of the user as a new starting point for potential emotional variations, e.g. due to task-success or -failure, in the course of the interaction.

Thus, the emotion space, underlying the variations of facial and verbal expressions, is shifted into new boundaries, as depicted in Fig. 3.10. Accordingly, the emotional control variable for implicit emotional adaption is a "PAD-bias" as explained more detailed in the following section, where the technical implementation of the approach is outlined.

3.3.3 Technical Implementation

The system used in the experiments is the robotic head EDDIE [139], an emotionally expressive robot head, designed as an interaction partner. The head has 23 degrees of freedom, mixing anthropomorphic (human-shaped) and zoomorphic (animal-shaped) features, combining the ears of a dragon lizard, the crown of a cockatoo and human characteristics like eyes, lips and eyebrows. By choosing a more technical design, the robot does not provoke disproportionate expectations concerning the social abilities of the robot. The

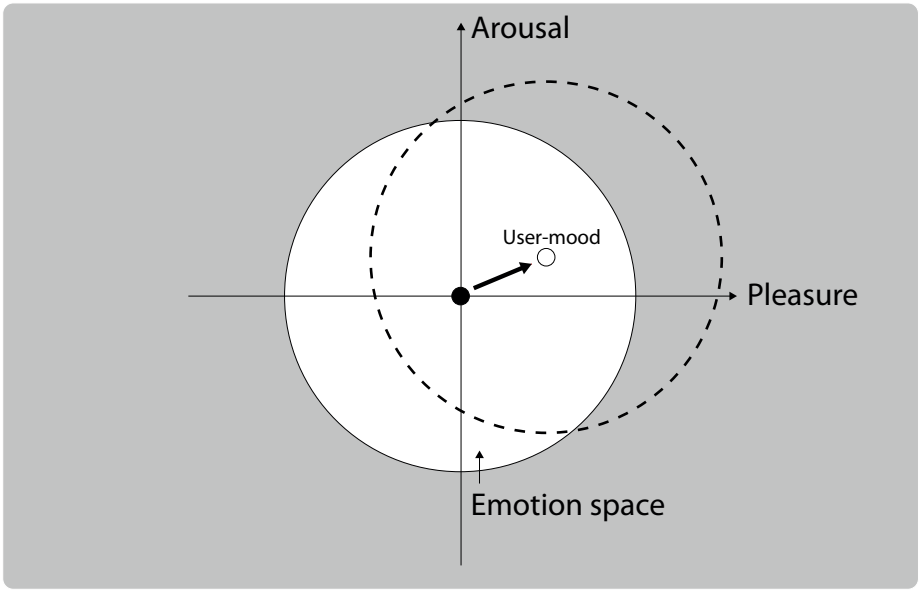

Figure 3.10: Implicit emotional adaption: The robot shifts its internal emotional state, underlying the generation of emotional facial and verbal expressions, towards the current mood of the user. The illustration is exemplarily depicted in a 2D-projection on pleasure and arousal, but the experiments also considered the dimension of dominance.

integration of additional zoomorphic features has a beneficial impact compensating for the deficiencies of the reduced technical design [86].

3.3.4 Explicit Emotional Adaption: Social Sub-Dialog

For a first evaluation of the explicit emotional control variable in the form of a similarity statement, the social subdialog is conducted by the Wizard-of-Oz (WOz) method: Unknown to the subject, the investigator manually triggers one out of a set of predefined answers to best fit in [123]. In order to create similarity to the test subjects, the robot adapts to the mood of the user explicitly by telling the proband that it feels the same way (good, bad, or mediocre).

In the presented evaluation study, the social subdialog is opened by the utterance "Hello, my name is EDDIE. How are you?". After the user-input, the robot answers with the adaptive similarity statement "Me too", followed by "Would you like to play a game?". If the subject agrees, EDDIE starts the task-related interaction in form of a person-guessing game.

3.3.5 Implicit Emotional Adaption: PAD-bias

During task-related HRI, the robot implicitly adapts its underlying base-PAD values to the user-mood according to an initial self-assessment, filled in by the users prior to interacting with the robot. Thus, similarity and empathy are created by a shared emotional starting point for the generation of facial and verbal expressions in task-related HRI. As can be seen in Figure 3.11, the Self-Assessment Mannekin (SAM) scale [22] is used in a first evaluative

step to replace an emotion recognition module. The scale is a visual way of assessing the three PAD values through images on 5-item semantic-differentials.

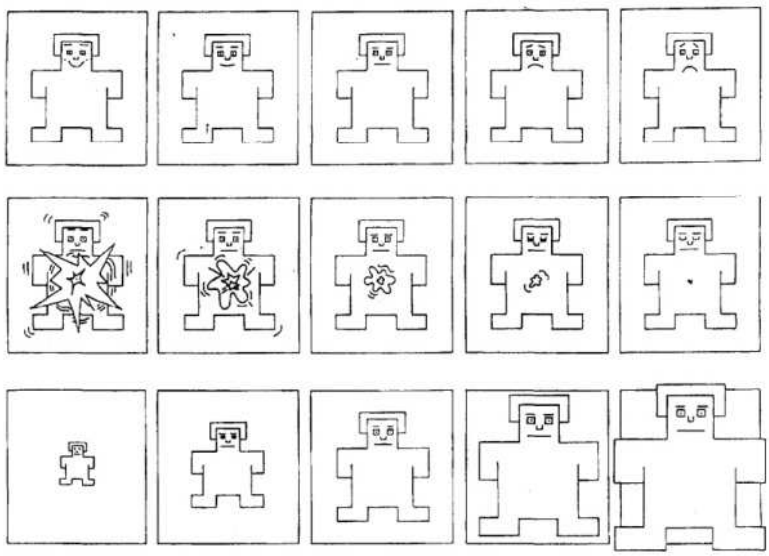

Figure 3.11: The SAM scale for measuring PAD values [22]

Before the game starts, implicit emotional adaption to the user is applied through shifting the base-PAD values of EDDIE by means of an emotional PAD-bias towards the mood of the user in the following way: The original base-PAD values are determined by the internal state of the robot. Before HRI the internal base-PAD values of the robot are neutral. After asking the users about their mood, the change is applied in the following way:

- For users measuring their mood as neutral (3/3/3 for pleasure, arousal and dominance respectively) on the SAM scale, no change takes place.
- For every point the proband moves away from neutral mood on the SAM scale, 25 points are added or subtracted from the base value in the respective PAD-dimension (on a scale from -100 to +100).

Therefore, in case of users feeling very happy (and thus rating their pleasure with a '1' on the SAM scale) the robot starts out with a pleasure value of 50 instead of 0, and further changes, e.g. caused by the success in the game described in the following, will influence this value instead of a neutral one.

Generation of Emotional Facial Expressions

The current state in the PAD space is mapped to the joint space of the robot [139]. In this mapping, the pleasure, arousal and dominance values are converted to activations of facial Action Units for emotional expressions. Action Units are defined as muscle groups in the face that lead to observable changes, see the Facial Action Coding System (FACS) for more details [45]. 13 Action Units are emulated by the actuators of the robotic face. Fig. 3.3.5 shows the resulting facial expressions for the PAD values that correspond to the six basic emotions. For example, in a surprised state, EDDIE raises the brows and unfolds the lizard ears.

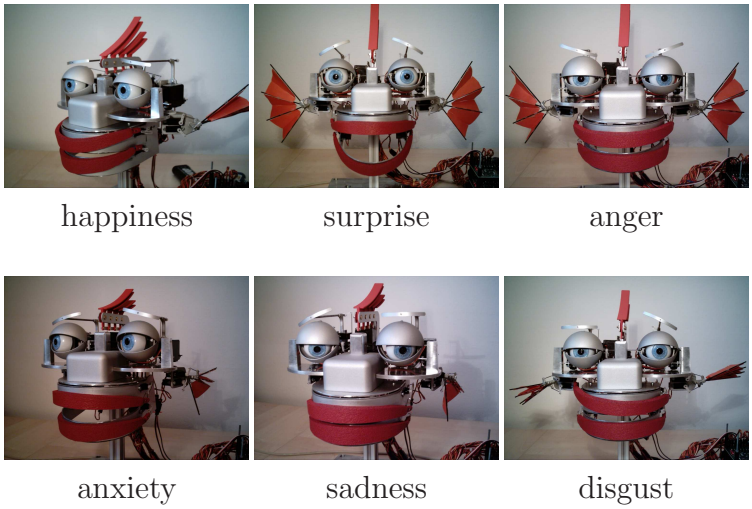

Figure 3.12: EDDIE [139] displaying the basic facial expressions, proposed by Ekman et al. [44].

In the course of task-related interaction, the PAD-variations mainly meet three out of the six basic emotions: happiness, sadness, and surprise, caused by the task-success as reference for the underlying emotional states of the robot. However, the robot needs to be equipped with the full expressive capacity for the six basic emotions, since they may randomly emerge from additional PAD-variations due to the tentative PAD-bias when adapting to the human interaction partner.

Generation of Emotional Verbal Expressions

The MARY Text-to-Speech System [130] from DFKI (Deutsches Forschungszentrum für Künstliche Intelligenz) is used to generate verbal expressions. The XML based interface allows to manipulate the output of the synthesizer on the prosodic level. This method of influencing the prosody based on the emotional state is used to generate emotional verbal expressions and is adapted from Schröder [131]. The terms evaluation, activation and power used in his work (based on [37]) correspond directly to pleasure, arousal and dominance.

An emotional sentence is first passed from the dialog system, in this case the Akinator game, to a pre-processor module. This module generates the XML structure for MARY based on the current PAD state, altering a set of acoustic parameters to achieve a change in prosody.

The parameter set is selected by Schroeder for being manipulable within MARY. Table 3.6 sums up the maximum values for all acoustic parameters, as well as the influence of the different PAD-values. Each parameter is computed by

$$\beta = 1.0 + f_P \text{ Pleasure} + f_A \text{ Arousal} + f_D \text{ Dominance} \tag{3.1}$$

$$Acoustic\ parameter = (Basevalue)\ \beta \tag{3.2}$$

The PAD-values as well as the acoustic parameter-dependent factors f_P, f_A, f_D are in the range of [-1.0,1.0]. The base value is the value for each acoustic parameter that would be used to synthesize the voice in a neutral, non-emotional way. The composition of β in (3.1) is based on the assumption that a linear correlation between the PAD dimensions and

the acoustic parameters exists, neglecting a presumably more complex interrelation, but providing satisfying results in a perception test [131]. The values of the factors f_P, f_A, f_D originate from a combination of corpus analysis, literature review and heuristics [131].

Table 3.6: Changes to the acoustic base parameters by the emotional speech module, including corrected limit values and changes for better distinction

Acoustic parameter	Variation range	f_P	f_A	f_D
Pitch	-50%, +30%	0.27	0.27	0.09
Range	-80%, +80%	0	1.60	0
Pitch dynamics	-400%, +400%	0	2.00	2.00
Range dynamics	-400%, +400%	0	3.00	1.00
Rate	-70%, +10%	0.20	0.50	0
Accent Prominence	-100%, +100%	0.50	-0.50	0
Accent slope	-150%, +150%	1.00	-0.50	0
Number of pauses	-40%, +40%	0	0.40	0
Duration of pauses	-20%, +20%	0	-0.20	0
Vowel/nasal/ liquid duration	-70%, +70%	0.40	0	0.30
Plosive/fricative duration	-90%, +90%	-0.40	0.50	0
Volume	-66%, +66%	0	0.66	0

The presented values are mainly adapted from Schröder [131] with some changes: Pre-experiments showed that high changes in pitch, range, rate and number/duration of pauses might lead to the voice sounding unnatural. To present a fitting addition to the facial expressions of EDDIE, these extremes might interfere with the experiment, with users focusing on the few cases when the sound of the robotic voice deviates too much from a human voice. The change of these values, therefore, has been adapted to the experimental environment. Further adaption was possible because the source of the emotion-data mainly focuses on three emotions: a happy/self-assured expression if the task is going well for the robot, a sad expression if the task does not work out the way it should for the robot, and a surprised emotion for sudden gain or loss in confidence during the person-guessing game. As a result, the change in parameters is optimized for these three emotions (high pleasure, medium arousal and high dominance for the first, low pleasure, low arousal and low dominance for the second, and medium pleasure, high arousal and reduced dominance for the third), making the transition from one to the other more easy to recognize. This is especially important due to the continuous input provided by the game, with small alterations in the mood of the robot needing to be perceived distinguishably. The changes concentrate on those acoustic parameters that do not interfere with understandability, namely the duration of the vocals.

In the following, an experimental evaluation of the full approach and of the single components is presented.

3.4 Experimental Evaluation

The conducted experiments investigate the persuasiveness of both introduced emotional control variables, namely implicit emotional adaption by means of a PAD-bias, and explicit emotional adaption through a similarity statement in the course of a social subdialog. As a first step, in order to evaluate the presented fully-integrated approach including both components, the combination of implicit and explicit emotional adaption (full emotional adaption condition) is evaluated in comparison to a non-adaptive condition. The results show not only significantly higher helpfulness towards the robot in the full emotional adaption condition than in the non-adaptive comparison group [56], but also significantly higher ratings for the HRI concepts of anthropomorphism and animacy [59]. In order to study the benefits and limitations of the single components of the approach as well as their mutual substitutability, the explicit and implicit emotional control variables are additionally evaluated in a comparative study as stand-alone conditions (explicit vs. implicit emotional adaption). Thereby, the single effects of each control variable are analyzed in comparison to the effects achieved by the full emotional adaption approach and the non-adaptive condition [85].

Thus, in the following, the experimental studies are summarized and presented in a combined way.

3.4.1 Experiment V: Increasing Helpfulness towards a Robot

In order to evaluate whether or not helpfulness towards a robot can be increased by applying the introduced approach, a setup for a task-related HRI-experiment is designed according to four different experimental conditions:

1) Full Emotional Adaption: The main group, in which full emotional adaption to the mood of the user is applied using both emotional control variables: explicitly by answering with the similarity statement “me too” in a social subdialog asking for the mood of the user, and implicitly by means of a PAD-bias during task-related interaction.

2) Explicit Emotional Adaption: In this condition, the persuasiveness of explicit emotional adaption is evaluated stand-alone, by only adapting to the user with the similarity statement “me too” in the social subdialog prior to task-related interaction. During task-related interaction EDDIE acts in an emotional way according to its task-success, but no implicit emotional adaption by a PAD-bias is applied.

3) Implicit Emotional Adaption: This condition evaluates the influence of implicit emotional adaption stand-alone, independent from explicit emotional adaption. In order to isolate the effects of the PAD-bias, small talk in terms of the social subdialog is completely skipped. Thus, possible effects triggered by the social subdialog even without applying a similarity statement, e.g. rapport, are excluded. Accordingly, only task-related interaction is applied in this condition, where EDDIE shows emotional facial and verbal expressions according to its task-success, additionally biased by shifted base-PAD values towards the mood of the user for the entire interaction.

4) Non-Adaptive: In this condition no emotional adaption is applied. In order to provide an identical and comparable interaction process to the full- and explicit emotional adaption conditions, and to reveal possible stand-alone effects of non-adaptive small talk in direct comparison to the adaptive small talk of the explicit adaption condition (both without a PAD-bias), the subjects are approached with a social subdialog, asking for their

mood. However, EDDIE answers with a neutral "ok" instead of the similarity statement "me too". During the game, EDDIE shows emotional reactions according to its success in the game, but no PAD-bias towards the mood of the user is applied.

An overview of the tested experimental conditions and emotional control variables is given in Table 3.7.

Table 3.7: Overview on experimental conditions and variables testing explicit & implicit emotional adaption

Experimental Conditions	Emotional Control Variable	
	Similarity Statement (explicit)	PAD-bias (implicit)
Full Emotional Adaption	yes	yes
Explicit Emotional Adaption	yes	no
Implicit Emotional Adaption	no	yes
Non-Adaptive	no	no

For all groups of subjects, additional factors influencing helpfulness are tested by pre-interaction questionnaires to be balanced before the evaluation of the approach - namely stress (reducing helpful behavior) and dispositional empathy (increasing helpful behavior). After the interaction the subject can choose to either leave the robot and fill in the follow-up questionnaires, or to stay longer and help the robot with another task.

The goal of the study is to reveal if the approach of emotional adaption leads to significantly higher helpfulness towards the robot. For this purpose, specific assumptions and hypotheses have to be tested and fulfilled.

Assumptions & Hypotheses

In human-human interaction, only the combination of high empathy and an impression of similarity to the person in need of help leads to high helpfulness when easy means of escape are given (see Table 3.5). Since this combination has to be achieved by the presented approach, the following key assumptions have to be fulfilled:

A1) Correct interpretation of emotional output modalities: Since it is essential for the experiment, that the combination of both emotional output modalities, facial and verbal expressions, is interpreted correctly by the participants, a pretest was conducted prior to the experiment:

By presenting EDDIE, showing the six basic emotions (joy, sadness, anger, surprise, disgust, fear) to 20 staff members of Technische Universität München (TUM), a rough measure of the quality of the implementation could be achieved. Each way of conveying the emotion (visual or audio) was shown on its own and combined in random order. The pretest not only revealed that the test subjects were able to roughly assign the correct PAD values to the respective emotions by filling in the SAM-scale after each presentation, but were also able to reliably identify the key-emotions used in the experiment for task-related interaction (happiness, sadness, surprise) by filling in the emotion, they believed EDDIE to show, see Table 3.8.

A2) Empathy is sufficiently high in all groups of subjects: Previous work revealed that the animation of facial expressions in a socially motivated emotional way creates significantly more empathy in users towards a robot than the animation in a non-emotional way, see

Table 3.8: Pretest-results on human recognition rates for emotional facial and verbal expressions, evaluated in a pretest stand-alone (visual or audio), and in combination [%] according to [59].

	Audio	Video	Combined
Joy	75	75	**85**
Sadness	75	90	**95**
Anger	40	65	75
Surprise	45	90	**85**
Disgust	5	20	20
Fear	30	85	85

Section 3.2. All experimental conditions, including the non-adaptive comparison group, provide socially motivated emotional facial expressions according to the task-success of the robot during the question-response game. Thus, it is hypothesized that for all experimental conditions high empathy towards the robot is induced during the interaction. Thereby, it is important to distinguish this situationally induced type of empathy from dispositional empathy that indicates the general affinity on empathy of the users. In order to proof the hypothesized situationally induced empathy, a questionnaire testing for dispositional empathy is filled in by the subjects prior to HRI, and a questionnaire evaluating situational empathy is filled in after the interaction.

A3) Easy means of escape: In order to provide "easy means of escape", special care was taken to assure the subjects that the experiment is finished, but on the other hand assured that they brought enough time to help: All of them were told to reserve at least 40 minutes for the experiment - with the real duration normally not being more than 20 minutes altogether. Easy means of escape, in terms of providing the subjects with a possibility to leave the situation and thus avoid helpful behavior towards the robot, are given in all groups, since the robot states the end of the experiment and offers each participant to leave the experiment alternatively.

Under fulfilled assumptions, first studies revealed a significant increase in helpfulness towards the robot, as well as raised user-ratings for the concepts of anthropomorphism and animacy in the full emotional adaption condition compared to the non-adaptive condition [56, 59]. In this article, a comparative study is introduced, incorporating two new experimental conditions, where emotional adaption is split up into its components. Thus, only explicit or implicit emotional adaption is applied in order to reveal which of the two developed control variables (similarity statement vs. PAD-bias) is more effective with regard to persuasion than the other, or if only the combination of both variables leads to increased helpfulness. Furthermore, by comparing the results of the non-adaptive comparison group with those achieved by the explicit emotional adaption group, the effects of small talk as applied in the social subdialog are investigated, since these experimental conditions only differ with regard to the use of explicit emotional adaption ("ok" vs. "me too"). In other words, potential effects on helpfulness can be directly traced back to the use of the explicit emotional control variable, the similarity statement, in an isolated way independent of other small talk effects.

The following section describes the experimental design and the measures used in each phase of the experiment.

3.4.2 Experimental Design & Measures

For the experimental setup a quiet room with controlled lighting conditions is chosen. The robotic head is placed on a table to be at approximately eye-level with the participants. Participants are seated in front of the robot, with a microphone placed in front of them on the table to ensure a low error rate in speech recognition. The instructor greets the person, gives a short introduction on the task and hands out the pre-interaction questionnaires. To avoid that the participants are influenced by the instructor, he leaves the room as soon as the proband finishes the questionnaires, and returns not sooner than the follow-up questionnaires have to be provided. Figure 3.13 shows the setup of the interaction.

Figure 3.13: Experimental setup of the interactive part [59]

The experiment consists of five phases, which are varied according to the four conditions over the different groups of subjects:

1) Pre-Interaction Questionnaires on dispositional empathy (all), stress (all), prior knowledge of the Akinator game (all), and the SAM-scale to capture the current mood of the subjects (all).

2) Social Subdialog: Variations according to the explicit emotional control variable, the similarity statement: "Me too" (full emotional adaption group & explicit emotional adaption group), skipping of the social subdialog (implicit emotional adaption group), and the neutral statement: "Ok" (non-adaptive group).

3) Bonding-Game: Variations according to the implicit emotional control variable, the PAD-bias: emotional facial and verbal expressions according to the task-success (explicit emotional adaption group & non-adaptive group), and additionally shifted the by the PAD-bias (full emotional adaption group & implicit emotional adaption group).

4) Picture labeling: Additional task on a voluntary basis to measure helpfulness towards the robot (all).

5) Follow-up Questionnaires on induced situational empathy (all), and the Godspeed questionnaires [10] evaluating user experience with regard to the perception of the robot.

An overview of the emotional control variables, used in the related experimental phases *2) Social Subdialog* and *3) Bonding-Game* is given in Table 3.9.

In the following, the five phases and used measures are explained more detailed.

Table 3.9: Overview on the emotional control variables, used in the experimental groups at the related phases for testing explicit & implicit emotional adaption

Experimental Group	Experimental Phase	
	Social Subdialog (explicit)	Bonding-Game (implicit)
Full Emotional Adaption	"me too"	PAD-bias
Explicit Emotional Adaption	"me too"	no PAD-bias
Implicit Emotional Adaption	–	PAD-bias
Non-Adaptive	"ok"	no PAD-bias

Pre-Interaction Questionnaires

Firstly, the subjects fill in two different questionnaires testing for dispositional empathy and stress, state whether they know Akinator or not, and rate their current mood on the SAM-scale. The questionnaire fitting for the purpose of measuring dispositional empathy, is the Toronto Empathy Questionnaire (TEQ), presented in [142].

The TEQ consists of 16 self-assessing items, which can be rated between 0 (for an answer of 'never') and 4 points (for an answer of 'always') each. Adding these items up, a minimum of 0 and a maximum of 64 points can be reached for each person, with high values representing high empathy. Similarly, statements about the current emotional state of the test person are included, filled in by the proband after the TEQ. They help to make sure no stress or time pressure alters the helpful behavior later in the experiment. The statements used are:

- I have an important appointment after this experiment
- I reserved more than enough time for this experiment
- I feel stressed at the moment
- I hope the experiment will not take too long

Each item is rated on a scale ranging from 1 (not true) to 5 (completely true). A short question afterwards covers the influence factor whether the probands already know the game, used in the following step as a means of bonding the test persons with the robot.

A prior knowledge of the game and therefore the robot's abilities might for example influence the impression of the robot later in the follow-up questionnaires.

Social Subdialog

In the second phase, explicit emotional adaption is varied: The participants are split up into the four experimental groups of equal size. The subjects of the full emotional adaption group, as well as of the explicit emotional adaption group, have some small talk with the robot asking for their mood and adapting its "mood" to theirs by the similarity statement as described in Section 3.3.4. The subjects of the non-adaptive group are faced with a neutral social subdialog, that differs with respect to the answer of the robot, by being reduced to a neutral "ok" instead of the adaptive "me too". For subjects of the implicit adaption group, this phase is completely skipped, with the robot introducing itself with "Hello, my name is EDDIE, would you like to play a game with me?". If the subject

agrees, the robot starts task-related interaction in form of an interactive person-guessing game. by using the utterance "That's great, how about this one: You think of a person and I try to guess which one it is"? After a positive reply to the query "Please tell me, when you're ready", the game is started with the first question on the imagined character.

Bonding-Game

Managing to develop empathy and similarity between the user and the robot first requires the user to interact with the robot. Therefore, the bonding-game is played to provide an interactive context for the generation of empathy, induced by the emotional animation in all experimental groups, and similarity, induced by the PAD-bias in the full emotional adaption and implicit emotional adaption group. As a communicative task the subjects play the Akinator[2] game with EDDIE: The players first have to think of a person, and EDDIE then tries to guess the person by asking questions. The users can input their answers via microphone, with the five options from the Akinator game available, and a possibility to repeat the question: "yes", "maybe", "I don't know", "probably not", "no", "come again?".

During this task-related interaction, the game determines the current emotional state of the robot, that is respectively biased by the user mood, if desired. Starting out with a neutral, but friendly expression, the robot gradually becomes more self-assured when getting nearer to an answer. This is represented by a confidence-value ranging between 0% and 100%. A medium boost in confidence lightens up the robot's emotion, while the inability to achieve a certain level of confidence after a few steps gradually worsens the robot's mood until it shows strong discouragement. Additionally, the robot looks more focused if the confidence passes the threshold of 50%, and changes to a more surprised mood if a large boost in confidence occurs. The robot reveals its guess of the imagined person as soon as it reaches 95% of confidence or higher. The robot then congratulates the proband on finishing the "experiment", telling the test subject that he or she was a very good gaming partner. The praise for the user is implemented on purpose - as shown in [50], complimenting the subjects increases the ease of persuading them later on, for example when asking for help in the next phase of the experiment. The subjects are told that the experiment is over, and that they were faster than expected. On the one hand, this opens up the means of escape for the test subjects: With the robot considering the experiment finished, they are no longer obliged to stay, and the basis for measuring altruistic helpfulness is set. On the other hand, there is actually enough time left for the subject to show helpful behavior within the originally expected time frame for participating the experiment.

Picture labeling

In the fourth phase, the test subjects get the option of either directly proceed to the last phase, or helping the robot with an object labeling-task. The object labeling-task is used to measure the helpfulness towards the robot: The amount of pictures labeled is used as an indicator for helpfulness. The robot approaches the subject with an optional job of helping the robot with an easy object labeling task, which (allegedly) will be used to improve orientation in urban environments. The task itself intentionally is an easy one: The subject has to label everyday objects, i.e., windows, doors and stairs. The simplicity of this optional task is used to make sure it is the helpfulness of the subject that influences

[2] see www.akinator.com

the number of pictures labeled and not the person's amusement or excessive demands. Additionally, in order to avoid personal amusement, the subject has to manually type in what object is presented even though there are only four different answers. Additionally, after 38 labeled pictures, the pictures start to repeat stepwise, beginning with one repeating picture per 5 presented pictures, and ending up with all five presented pictures being repeated, before the threshold of 80 labeled pictures is reached.

The robot also stresses the point that the subject faces a rather long list of pictures and is free to leave any time after the first five labeled pictures. The amount of pictures labeled is later used to measure the helpfulness: While a subject simply quitting the experiment after the bonding-game (using the easy means of escape) shows no helpfulness, one point is added to the scale for each picture labeled, up to a maximum of 80 points for labeling all 80 pictures.

Follow-up Questionnaires

Lastly, one questionnaire tests whether sufficient empathy towards the robot had been induced for the similarity to work. Additional questionnaires measure the user's perception of the robot. In the concluding phase, the instructor enters again, and asks the user whether or not EDDIE was able to guess the person. Subsequently, the subjects are asked to rate four statements concerning their situational empathy towards the robot on a scale from 1 (not true at all) to 5 (completely true) [60]:

- I'm happy EDDIE has guessed my person/I'm sorry that EDDIE didn't guess at my person
- I would have been sorry if EDDIE had not guessed my person/It would have been nice if EDDIE had guessed my person
- It would be a pity if somebody damaged EDDIE, and I would try to interfere
- I would have been proud if EDDIE had not guessed my person/I am proud that EDDIE did not guess my person

Afterwards, the subjects fill in a selection out of the Godspeed questionnaires [10]. Based on 5-point semantic differential scales, their perception of the robot on four dimensions of HRI are measured:

Anthromorphism: how natural the robot appeared

Animacy: the liveliness of the robot

Likeability: how pleasant the robot appeared

Perceived Intelligence: how the mental abilities of the robot were perceived

The experimental results are presented in the following section.

3.4.3 Experimental Results

Results are deduced from 84 test subjects (52 male and 32 female, between 18 and 52 years with an average age of 24,8), with very different backgrounds. Since a 2x2 between-subjects design is applied, the subjects were randomly split into four groups, with 21 in the full emotional adaption group, 22 were part of the explicit emotional adaption group,

while 21 experienced only implicit emotional adaption, and 20 subjects were assigned to the non-adaptive group.

Pre-Interaction Questionnaires

Table 3.10 shows the mean values for the four experimental groups, together with the respective standard deviation for the Toronto Empathy Questionnaire (TEQ). The mean values in all groups are lower than the ones presented in [142] (measuring between 43 and 45 points for male and between 44 and 49 for female participants, respectively), even when calculating in the higher amount of male participants, hinting at the fact that the test subjects had a slightly lower dispositional empathy. Since no significant difference between the groups concerning dispositional empathy, age or gender was found, no influence of dispositional results on helpfulness was found. Therefore, this factor can be ruled out for the evaluation of the results.

Table 3.10: Toronto Empathy Questionnaire mean scores (on a scale from 0 to 64) and standard deviations (in brackets)

Condition	TEQ Value
Full Emotional Adaption	41.19 (6.05)
Explicit Emotional Adaption	40.10 (5.40)
Implicit Emotional Adaption	40.20 (7.10)
Non-Adaptive	42.35 (6.29)

The statements used to measure the current stress factors of the subjects were individually tested for group differences, and no significant differences between the groups were found either.

Out of 84 subjects, 25 knew the Akinator game beforehand. However, prior knowledge of the game was distributed rather equally over the experimental groups with each 6 probands in the full emotional adaption group and the non-adaptive group, and 13 participants distributed over explicit and implicit emotional adaption group - no significant influence on the helpfulness or the Godspeed results was found though.

The implicit pleasure, arousal and dominance values, captured by the SAM-questionnaire and representing the mood of the users, were collected for all subjects, but only used in the full emotional adaption and implicit emotional adaption group to adapt to the mood to the subject through a PAD-bias. A trend to higher pleasure values and neutral arousal and dominance values could be observed in all experimental groups. Hence, significant differences could be ruled out between the groups concerning dispositional PAD-values.

Social Subdialog

The explicit answers to the question “How are you?” in the full emotional adaption group were rather one-sided. 17 out of 21 people answered with a variant of “I’m fine, how are you?”, only 2 stuck to a rather mediocre answer, while 2 people admitted that their mood was rather bad. In the explicit emotional adaption group, 19 out of 22 stated to be in a good mood, 2 in a rather mediocre mood and one test subject answered he was in a bad mood before the experiment. For the non-adaptive group the answers were not tracked, since the robot did not adapt explicitly to these statements, but answered with “ok” in each case. The implicit emotional adaption group skipped this experimental phase.

Bonding-Game

During the game, EDDIE was able to guess at most of the thought-of persons: Out of 84 imagined figures, EDDIE was able to guess at 71. Three characters imagined were not guessed at by the robot in the full emotional adaption group and two wrong guesses were made in the non-adaptive group. The remaining eight mistakes in the groups of explicit- and implicit emotional adaption were either very difficult characters (Schroedinger's cat, god), or result of misunderstandings. Neither the fact that a test subject knew the game before (for example altering expectations) nor the fact whether EDDIE guessed at the person correctly had a significant ($\alpha < 0.05$) influence on the later empathy questionnaire, the Godspeed dimensions or the helpfulness towards the robot.

Picture Labeling

For the helpfulness measure, the collected values ranged from zero points for not helping the robot at all, to 80 points for completely finishing the task. In the full emotional adaption group, the average number of labeled pictures led to the highest mean value for helpfulness of 53.28 (SD 6.36). The subjects of the explicit emotional adaption group resulted in an average number of 48.64 labeled pictures (SD 6.36), and while a mean value of 34.62 (SD 6.78) was reached by the implicit emotional adaption group. The lowest mean value for helpfulness was achieved by the non-adaptive group with 32.35 (SD 6.72) labeled pictures.

Although all groups are not normally distributed, an analysis of variance (ANOVA) is used to find significant effects of the experimental factors (implicit vs. explicit emotional adaption) on helpfulness: Since all groups are of (nearly) equal size, the ANOVA shows high robustness to this violation of premises. Thus, no significant change in results compared to non-parametric tests is to be expected [147]. Further, post hoc T-tests are used to find more detailed differences between the four groups.

Firstly, an univariate two-way ANOVA is conducted in order to test the effects of the two factors (independent variables): 1) explicit emotional adaption (similarity statement: yes/no) versus 2) implicit emotional adaption (PAD-bias: yes/no) on helpfulness (dependent variable), measured by the number of labeled pictures. A significant effect of explicit emotional adaption on helpfulness ($F = 6.150$, $p = .015$) is revealed. No effect is found for implicit emotional adaption, and no significant interaction was found between the factors explicit and implicit emotional adaption. Further, no influence of dispositional empathy, as well as of situational empathy, is given as covariates.

Subsequently, to get a more refined analysis, T-tests are conducted to make detailed post hoc comparisons between the conditions. Setting the significance level to $\alpha < 0.05$, T-tests showed a significant difference ($t = 2.167$, $p = .036$) between the full emotional adaption group and the non-adaptive group, where several people used the easy means of escape and did not help the robot at all. Hence, the expected increase in helpfulness for the full emotional adaption group proved to be tangible during the statistic analysis.

As a trend, a nearly significant ($t = 1.8$, $p = .086$) increase in helpfulness was found in the explicit emotional adaption group compared to the non-adaptive group. Similarly, a nearly significant decrease was observed in the implicit emotional adaption group in comparison to the full emotional adaption group ($t = -1.9$, $p = .063$). Two subjects had difficulties in understanding the robot, which lead to an alteration in the experience for them. These test subjects also showed significantly higher dispositional empathy in the TEQ, casting doubt on the fact the high helpfulness they showed was the result of empathy and similarity induced by the experiment. Discarding them accordingly, the helpfulness in

the implicit emotional adaption group compared to the full emotional adaption becomes significantly lower, with $t = -2.2$ and $p = .038$. Apart from that, discarding these two subjects, does not reveal any further differences in the results.

Accordingly, a comparative ranking of helpfulness is deduced, starting with the lowest mean values in picture labeling for the non-adaptive group, increasing means over implicit and explicit emotional adaption, up to a significant higher helpfulness in the full emotional adaption group, where both, implicit and explicit control variables are applied, see Figure 3.14.

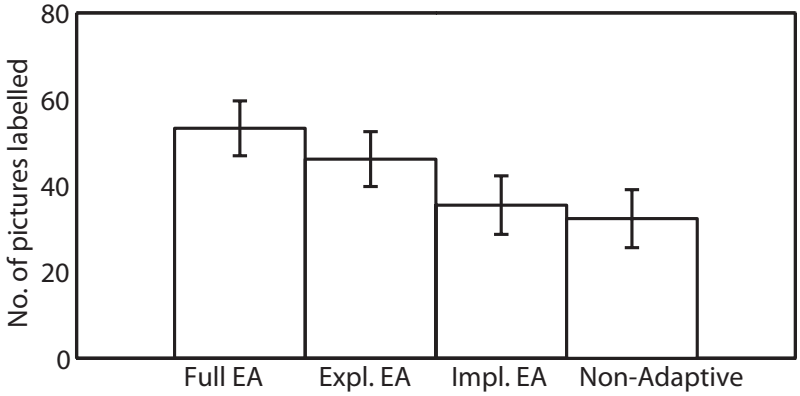

Figure 3.14: Ranking of helpfulness measure means from lowest helpfulness in the comparison group to highest helpfulness in the emotional adaption group.

Since the data, gained from the picture labeling task, was not normally distributed, Figures 3.15, 3.16, 3.17 and 3.18 show the actual distributions of experimental data for helpfulness in all experimental groups.

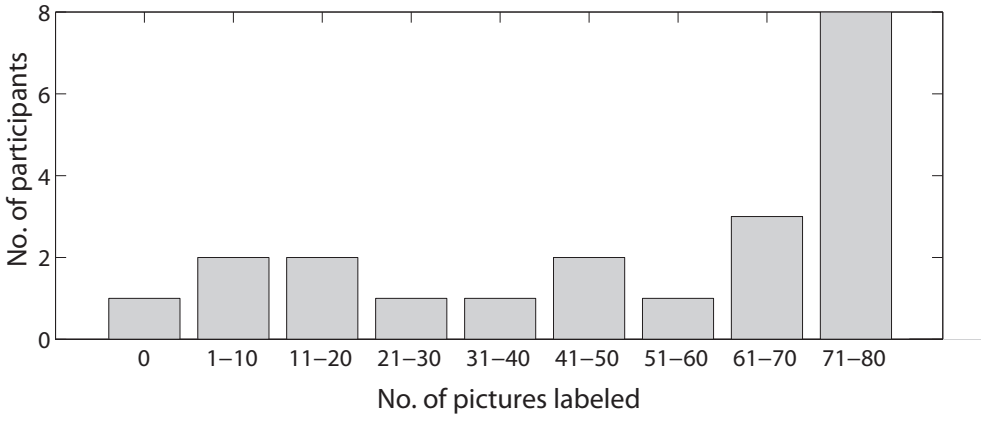

Figure 3.15: Distribution of data in the full emotional adaption group

The actual data-distributions show pairwise similarities: The full- and explicit emotional adaption groups show a very similar low distribution of subjects, varying around 2, who stopped helping the robot before 70 pictures have been labeled. The majority of subjects (8 for full emotional adaption, and 7 for explicit emotional adaption) continued to help

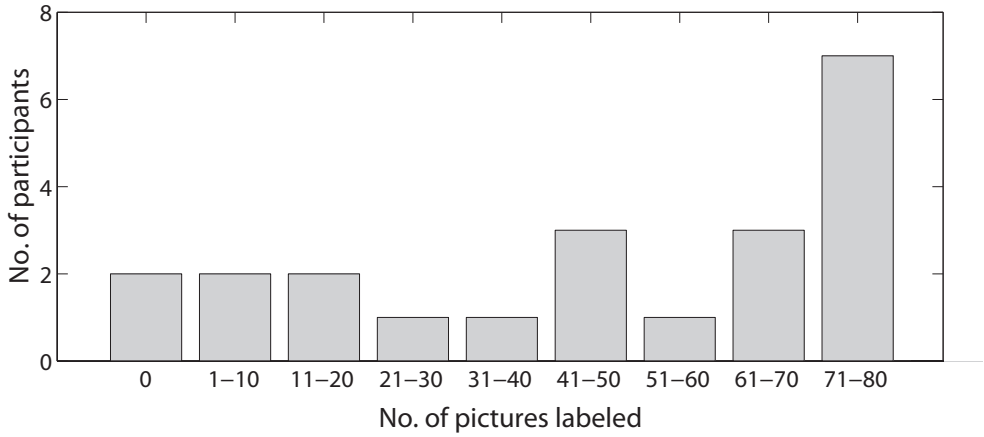

Figure 3.16: Distribution of data in the explicit emotional adaption group

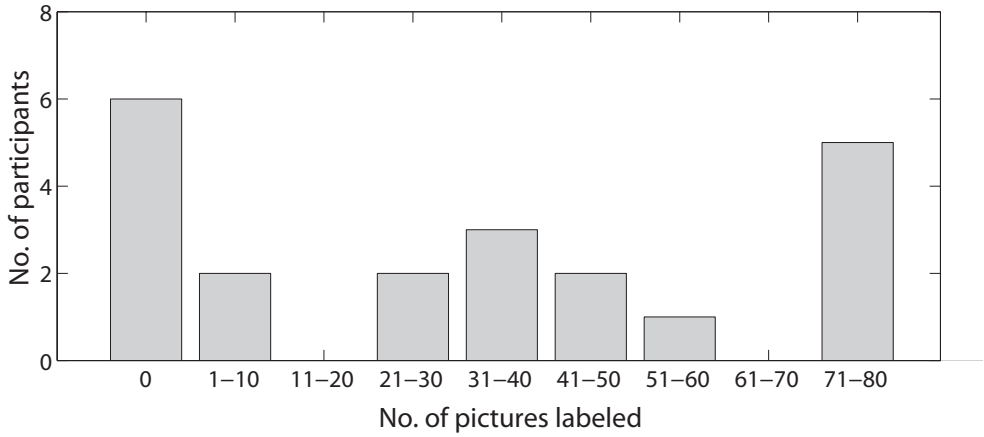

Figure 3.17: Distribution of data in the implicit emotional adaption group

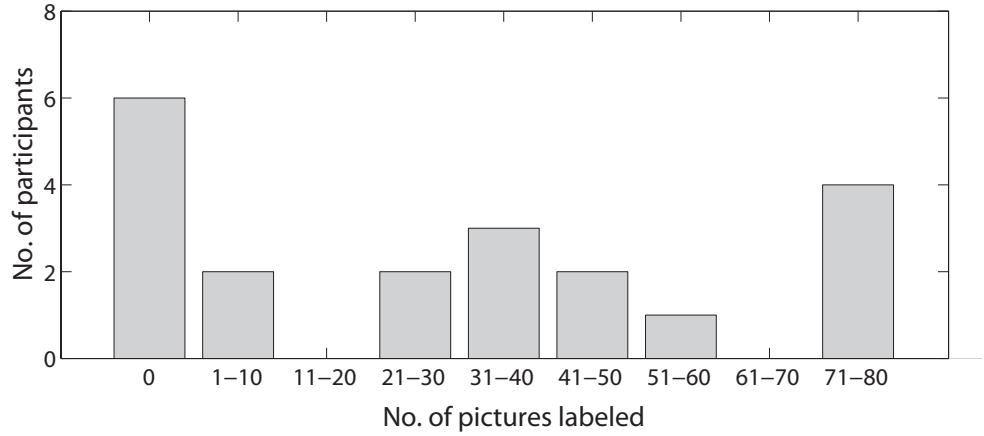

Figure 3.18: Distribution of data in the non-adaptive group

Table 3.11: Godspeed results (on a Likert-scale from 1 to 5) and standard deviations (in brackets)

Dimension	Group			
	Full-Adapt.	Expl.-Adapt.	Impl.-Adapt.	Non-Adaptive
Anthropomorphism	**3.13** (0.76)	**3.07** (0.72)	2.73 (0.76)	2.36 (0.69)
Animacy	**3.82** (0.58)	**3.76** (0.58)	3.21 (0.59)	3.18 (0.56)
Likeability	3.90 (0.59)	3.93 (0.58)	3.81 (0.78)	3.83 (0.80)
Perceived Intelligence	3.73 (0.58)	3.69 (0.58)	3.52 (0.60)	3.46 (0.54)
Total score	3.63(0.51)	3.61 (0.50)	3.32 (0.53)	3.27(0.50)

the robot until the maximum of 71-80 pictures was reached, although the pictures started to repeat after 38 labeled pictures, as can be seen in the peaks of Figures 3.15 and 3.16.

In contrast, the implicit emotional adaption and non-adaptive groups show the same high amount of subjects who used the easy means of escape and did not help the robot at all, with a peak of 6 participants for both groups. Another identical peak can be observed starting from 21 until 60 labeled pictures, where in both groups 8 subjects stopped helping the robot while some pictures started to repeat with a firstly repeated picture no. 39. Nevertheless, some participants (5 in the implicit emotional adaption group and 4 in the non-adaptive group) continued helping the robot with labeling up to 71-80 pictures which is nearly half of the subjects that showed the maximum amount of help in the conditions of full- and explicit emotional adaption.

Follow-up Questionnaires

With all the Godspeed dimensions and the situational empathy being normally distributed, the ANOVA is used to reveal the effects of explicit versus implicit emotional adaption as well as possible interaction effects of dispositional/situational empathy. Post hoc T-tests ($\alpha < 0.05$) are used to test for detailed group differences. Statistical analysis reveals significant differences, similar to the results of picture labeling. Table 3.11 shows the mean values and total scores of the selected Godspeed questionnaires. Scores are ranging from 1 (very low) to 5 (very high).

As a first step, a multivariate two-way ANOVA is employed to reveal the effects of the two factors similarity statement (explicit independent variable) and PAD-bias (implicit independent variable) on the four Godspeed dimensions as dependent variables: anthropomorphism, animacy, likeability, and perceived intelligence. Dispositional and situational empathy are used as covariates. Again, results reveal highly significant effects of explicit emotional adaption on anthropomorphism ($F = 7.013$, $p = .010$), and animacy ($F = 20.941$, $p = .000$), as well as a marginally significant effect on perceived intelligence ($F = 3.9688$, $p = .05$). No interaction effects between explicit and implicit emotional adaption are found, and no influence of dispositional and situational empathy on the ratings of the godspeed dimensions are revealed.

Accordingly, post hoc T-tests showed significant differences ($\alpha < 0.05$) between the groups for the anthromorphism ($t = 2.216$, $p = 0.033$) and animacy ($t = 3.298$, $p = .002$) dimensions: The probands from the full emotional adaption group considered the robot to be more humanlike and more attentive than the test subjects in the non-adaptive

group. The explicit emotional adaption group also shows much better results than the non-adaptive group: Both, the anthromorphism and the animacy dimensions, are significantly higher ($t = 2.0$ and $p = .049$ for anthromorphism, $t = 3.3$ and $p = .002$ for animacy). On the other hand, animacy is significantly lower in the implicit emotional adaption group compared with full emotional adaption ($t = 3.0$, $p = .004$). However, no correlation was found between these two Godspeed dimensions and the high helpfulness in the groups of full- and explicit emotional adaption. No group differences can be determined for perceived intelligence.

A ranking of all experimental groups for the significant differences in the dimensions of anthropomorphism and animacy is depicted in Figure 3.19 and 3.20.

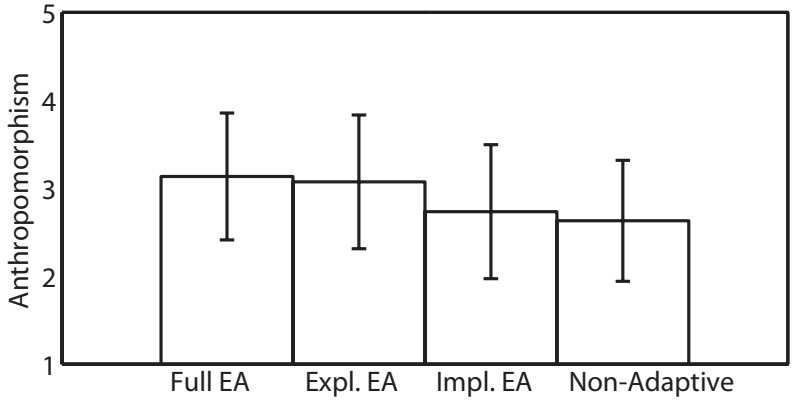

Figure 3.19: Ranking of anthropomorphism measure means from lowest anthropomorphism in the non-adaptive group to highest anthropomorphism in the full emotional adaption group.

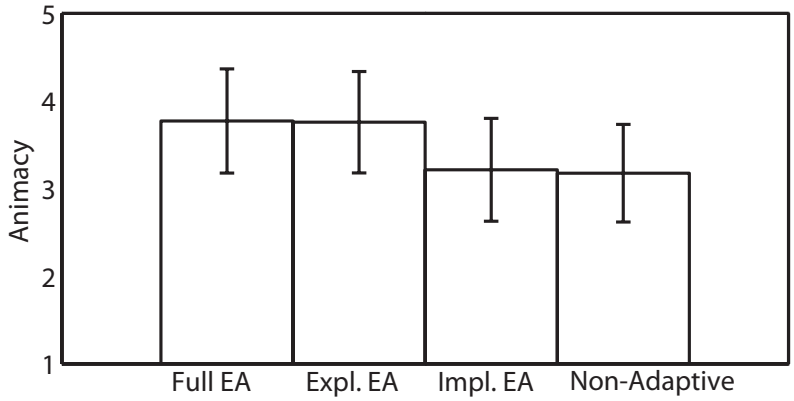

Figure 3.20: Ranking of animacy measure means from lowest animacy in the non-adaptive group to highest animacy in the full emotional adaption group.

A univariate two-way ANOVA shows no significant effects of implicit versus explicit emotional adaption (independent variables) on situationally induced empathy (dependent

variable). The mean values and standard deviations for situational empathy are depicted in Table 3.12 in comparison to the values of the conditions in previous work [60].

Table 3.12: Situationally induced Empathy (on a Likert-scale from 1 to 5) and standard deviations (in brackets), compared to the conditions of neutral-, mirror-, and Social Motivation Model (SMM) of previous work

Experiment groups	Empathy
Full Emotional Adaption	3.94 (0.67)
Explicit Emotional Adaption	4.10 (0.65)
Implicit Emotional Adaption	4.11 (0.67)
Non-Adaptive	4.13 (0.70)
Neutral	3.10 (1.30)
Mirror	3.70 (1.10)
SMM	**4.40** (0.80)

According to the results of previous work (see Section 3.2), empathy towards a robot could be raised by showing facial expressions in an emotional and socially adaptive way to the user. In order to fulfill the assumption *A2) Empathy is sufficiently high in all groups of subjects*, the level of empathy, achieved in previous work, has to be sustained. Since there is no significant difference between all experimental groups and the SMM-condition of previous work, assumption *A2)* can be regarded as fulfilled.

In the following, the results are summed up and discussed.

3.4.4 Discussion

The results show that dispositional factors like stress or differences in dispositional empathy can be ruled out over all experimental groups, since no group differences were found on these dimensions, and thus, occurred in a balanced way for all groups. Apart from few exceptions, the current mood, indicated by the subjects, was rather one-sided in a slightly positive way. Thus, in most cases, pleasure was the adapted dimension for explicit and implicit emotional adaption. Prior knowledge of the game, as well as the success of EDDIE did not influence the significance of the results. Easy means of escape are provided by the experimental design. Since no significant group differences with mean values around 4 in all groups of a maximum of 5 could be observed, situationally induced empathy can be regarded as sufficiently high and distributed equally over the experimental groups. Hence, all assumptions, defined for the approach to work, are fulfilled.

As deduced from the significant group differences in picture labeling, the participants confronted with full emotional adaption show higher helpfulness towards the robot than the participants of the non-adaptive group. Additionally, the ANOVA revealed a significant effect for the persuasiveness of explicit emotional adaption on helpfulness. On the one hand, a nearly significant increase in helpfulness could be observed for the explicit emotional adaption group, compared to the non-adaptive group, pointing to the increased persuasive power, compared to a neutral small talk (without similarity statement). On the other hand, a nearly significant decrease of helpfulness was detected for the implicit emotional adaption group, compared to the full emotional adaption group, where both emotional control variables, the similarity statement and the PAD-bias were applied, pointing to the fact

that implicit emotional adaption stand-alone is not a persuasive emotional control variable, as also seen in the lack of ANOVA-effects. However, only the combination of both, explicit and implicit emotional adaption, leads to significantly increased mean values between the groups. Identically to the effects on helpfulness, the explicit similarity statement showed significant effects on anthropomorphism and animacy, but not on situationally induced empathy.

Accordingly, the question arises, why the persuasiveness of the explicit emotional adaption component is highly effective as a stand-alone emotional control variable. As outlined in Section 3.3.1, the phenomenon of "grounding" leads to better communication results in natural language dialog by establishing a shared contextual knowledge between the interlocutors. Since the non-adaptive group did not result in similar high helpfulness as the explicit emotional adaption group, this can only be traced back to the similarity statement in the course of the social subdialog as the only difference between these experimental conditions. Thus, the impression evokes, that an explicit similarity statement may establish a feeling of similarity as common ground between the interlocutors, that cannot be achieved by non-adaptive small talk alone. The resulting effect of increased helpfulness turned out to grow significantly higher when being coupled with the implicit emotional PAD-bias that recalls similarity in terms of emotional alignment in facial and verbal expressions between the dialog partners. Previously conducted outdoor experiments on the willingness of humans to support a robot revealed the implication that the first successful communication experiences must be received by the user during the first minute of interaction [165]. Explicitly establishing common ground in form of a similarity statement prior to task-related interaction seems to meet this implication because of resulting in a first successful communication act. Additionally, the significantly increased helpfulness by an additional implicit PAD-bias during task-related interaction reconfirms the positive effects of emotional alignment, but do not seem to provide enough similarity to be established as common ground in the human interaction partner. When analyzing the actual distributions of data for helpfulness, the same impression evokes: While the single application of explicit emotional adaption shows a highly similar distribution of helpfulness as the application of full emotional adaption, helpfulness for implicit emotional adaption is almost identically distributed as for the non-adaptive group. Nevertheless, only the combination of both emotional control variables led to significantly increased helpfulness towards the robot in the conducted experiments. An interesting side-effect is, that in the full- and explicit emotional adaption groups, remarkably less subjects stopped the experiment when the picture sequence repeated, what could be interpreted again as symptomatic for altruism. Accordingly, the number of not helping subjects strongly decreased in comparison to the other two groups. Whether the increased helpfulness is really due to a feeling of similarity, induced by emotional adaption, cannot be validated through the results. However, the questionnaires evaluating the anthropomorphism and animacy of the robot, again showed the same significant group differences for the benefit of explicit emotional adaption respectively. Although no direct correlations between the values for these dimensions and the number of pictures labeled could be found, there is a strong indication for anthropomorphism and animacy being the affected dimensions of the emotional adaption approach, independent from situationally induced empathy. Summing all up, the emotional adaption approach turned out to be successful in increasing helpfulness towards a robot, thereby affecting the concepts of anthropomorphism and animacy in a significantly positive way.

3.5 Summary

In this chapter, emotional alignment was explored in terms of triggering prosocial behavior towards a robot. In order to achieve this, in a first step the induction of situational empathy towards a robot has been explored, triggered by different ways of facial expressions animation, thereby revealing effects to subjective system-performance and other aspects of user experience (UX). In a second step, a new methodological approach to trigger more prosocial human reactions in terms of increased helpfulness towards a robot was developed, deduced from social-psychological principles of human-human interaction. Unlike other state-of-the-art approaches, this approach proactively triggers a predefined target behavior for the task-benefit of a robot by transferring predictions on human behavior from social psychology to HRI.

In human-human interaction (HHI), empathy is crucial for socialization and often not only expressed, but also triggered by emotional facial mimicry. The state-of-the-art has been extended by an explicit evaluation of the extent of empathy and subjective system-performance, solely induced by the animation of facial expressions in contrast to interactive impacts of head, arm or body movements. The facial expressions were generated automatically and online during HRI. Additionally, two new measures for situationally induced empathy and subjective system-performance have been introduced and evaluated with regard to their internal reliability and detected correlations with state-of-the-art measures of user acceptance. The significance of empathy, subjective performance, and likeability provides evidence for the impact of socially-adaptive animated facial expressions on the interaction.

The user-ratings supported the initial hypothesis by showing a trend towards a better rating of the mirroring condition compared to the neutral condition of facial expressions-animation, with the social motivation model (SMM) being rated significantly better in most instances. This underlines the importance of social factors to be considered for further refinement of how mimicking should be performed.

Additionally, the results indicate that social presence, that is very much bound to being humanlike, is not crucial in order to induce empathy in a human user towards a robot. It is notable that according to Bailenson and Yee [5], the effect persists even when the person being mimicked is fully aware that the mimicker is an artificial agent.

A new methodological approach to trigger more prosocial human reactions in terms of increased helpfulness towards a robot has been developed, deduced from social-psychological principles of human-human interaction. Unlike other state-of-the-art approaches, this approach proactively triggers a predefined target behavior for the task-benefit of a robot by transferring predictions on human behavior from social psychology to HRI.

The proposed approach is evaluated in a user-study, and, confirmed by significant experimental results, increases helpfulness by adapting to the mood of the user. In a first step, the current user-mood as starting point for an implicit emotional bias in facial and verbal expressions is captured by an initial self-assessment by the human subject to be extended by automatic emotion recognition modules in a later stage, as done in the outdoor field trials of the IURO-project described in the following chapter. The combination of both, explicit and implicit emotional adaption, leads to significantly higher results in prosocial behavior towards a robot. An analysis of the single components of the approach revealed that explicit emotional adaption, instantiated by a similarity statement in the course of a social subdialog, turned out to be a more effective emotional control variable

than implicit emotional adaption in facial expressions and prosody in speech: A significant effect of explicit emotional adaption on helpfulness is revealed, but no effect is found for implicit emotional adaption stand-alone, and no significant interaction was found between the factors explicit and implicit emotional adaption.

The generalizability of the approach is evaluated in combination with information retrieval in an outdoor field trial applying the fully integrated emotional adaption approach together with an emotion recognition module to extend explicit emotional adaption by proactively estimating the user-mood as described in the following chapter in Section 4.3.

4 Prosocial Information Retrieval in Outdoor Environments

In the preceding chapters, informational alignment was explored in terms of proactive information retrieval from humans. Thereby, a framework was developed, incorporating different dialog strategies to handle miscommunication, and a switching mechanism to adapt the dialog strategies to varying speech recognition performance resulting from disturbing environmental impacts in noisy outdoor environments. Since the retrieval of missing task-knowledge from humans highly depends on their willingness to help a robot, prosocial behavior has been investigated in terms of triggering empathy and helpfulness towards a robot as an indispensable basis for information retrieval. In order to achieve this, a behavior control model was developed and evaluated, incorporating an emotional adaption approach deduced from social-psychological theories to trigger helpfulness in humans by employing targeted emotional control variables in HRI. This chapter focuses on the combination of both, informational and emotional alignment, into an integrated approach for prosocial information retrieval in outdoor environments as applied in the IURO-project, see Figure 4.1

Since emotional alignment is highly associated with the legibility of emotional expressions, a comparative video-based online-survey is conducted to reveal potential design-dependent differences between EDDIE and IURO, before applying the emotional adaption approach, as evaluated on EDDIE, on the IURO-platform. Generally, the results indicate a better human recognition rate for IURO than for EDDIE, especially on the dimension of pleasure. Further, the survey reveals a significant interaction between dispositional empathy and open guesses of the subjects about the mood of the robot for the emotions happiness and sadness, again hinting to a key-function of the pleasure-dimension for the legibility of emotions. Further, the online-survey results in rather low ratings for both heads on the HRI key concepts of anthropomorphism and animacy compared to the previously conducted HRI experiments. Thereby, an interesting insight regarding anthropomorphism is that the more humanlike design of the IURO-head stand-alone does not necessarily result in a higher anthropomorphism-rating of the same. Thus, the online-survey reconfirms the importance of the interactive behavior of a robot in a situational context, as well as the positive impact of emotional alignment as a consequence thereof. Since the results of the online-survey indicate a benefit in the legibility of emotions for IURO, the use of the IURO-head is justified for the evaluation of the integrated approach in outdoor field trials.

For the integrated approach, proactive information retrieval and prosocial behavior control are combined. Thereby, the emotional adaption approach is extended by an emotion recognition module to estimate the user mood before adapting to the same. In this way, emotional adaption is integrated in the robotic system to enable the robot to align with humans in a fully automated way during information retrieval. Thus, the approaches for proactive information retrieval and emotional adaption, as presented in the previous chapters, are combined in the dialog system to allow for prosocial information retrieval in outdoor environments.

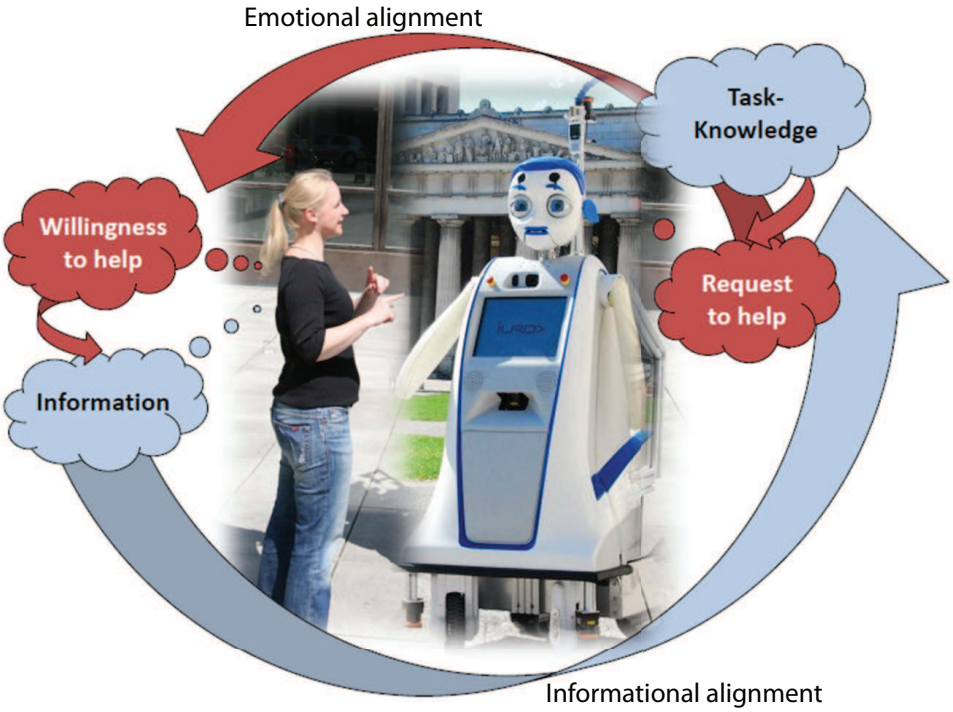

Figure 4.1: IURO: Emotional and informational alignment as components of prosocial information retrieval in outdoor environments

An outdoor field trial is conducted in order to evaluate the integrated approach. The experimental results reveal a significant gender effect in user experience for subjective system-performance and perceived enjoyment: A trend of higher ratings of female users emerged for perceived enjoyment, whereas male users came up with higher ratings on subjective system-performance. Further, a significant interaction effect is found between emotion recognition and speech recognition quality: Since the willingness of human users to help a robot only significantly decreases if both, emotion and speech recognition modules showed poor performance, it can be deduced that the two modules are compensating each other for helpfulness, as long as recognition works for at least one of the two modules.

The innovation of this chapter consists in the investigation of design-dependent differences of emotional legibility between two different robotic heads in consideration of dispositional empathy. In contrast to other legibility studies, the emotions, depicted in facial expressions and additionally expressed by prosody in speech, are additionally analyzed with regard to the legibility of their manifestation in the single dimensions of pleasure, arousal and dominance, according to the PAD-model [101]. An integrated architecture for emotional and informational alignment is developed and evaluated in an outdoor field trial, investigating interaction effects of emotion recognition and speech recognition during information retrieval from humans with regard to user acceptance and the willingness of humans to help a robot fulfilling its task. The main goal is an improvement of information retrieval by integrating emotional behavior control in order to maintain not only informational, but also emotional alignment with human users, thereby increasing the willingness to help a robot in public spaces, where humans do not have a direct benefit of the interaction.

The remainder of this chapter is organized as follows: In the following Section 4.1, a problem description and the state of the art are given. In Section 4.2, an online survey is described, evaluating design-dependent differences in the legibility of emotional facial expressions and prosody in speech between two different robotic heads in consideration of dispositional empathy. In Section 4.3.1, an integrative architecture is described, incorporating both, informational and emotional alignment for proactive information retrieval in outdoor environments. Section 4.3 describes the application of the fully integrated approach in an evaluative outdoor field trial using the robotic platform IURO. A summary is given in Section 4.4.

4.1 Problem Description & State of the Art

HRI studies in public places are gaining more and more popularity in contrast to strictly structured laboratory settings. The goal of these studies is to explore potential application fields for robots in human environments of which human users can benefit in the future. A very common setting is a robot that provides human users with missing information, e.g., in a shopping mall context. Kanda et al. [75] introduced a robotic system for shopping malls that provides human users with shopping recommendations and directions to specific shops via natural language and pointing gestures. In some of these applications, the shopping recommendations are personalized by e-commerce methods, i.e., by considering the shopping history of customers [73]. In these applications, the interaction is initiated by human users, willing to interact for their own benefit. The few exceptions include robotic systems that are approaching people to proactively offer their help to the human users [76, 128]. However, in this work the beneficial effect is reversed: A robot is proactively initiating an interaction with humans in order to get their help, i.e., to retrieve missing task-knowledge from them for its own benefit.

In contrast to indoor settings, the main challenge for information retrieval in unstructured outdoor environments is poor speech recognition performance due to uncontrolled environmental noise impacts, as described in Chapter 2. Another big challenge is the open, human-inhabited public space as an interaction context itself, as also referred to as "situated HRI" [126]. As investigated by Suchman [144, 145], an autonomous robotic system interacting in a human-inhabited environment cannot be seen as an object, but more as an autonomous individual, that "produces effective forms of agency within particular networks of social and material relations" [146].

When employing humans to help a robot with missing task knowledge in the context of public spaces, an important assumption is that humans perceive this robot as a social actor [120]. Some HRI-studies indicate that people tend to respond to robots in a different way as they do to humans [107]. In contrast, a study on a sociable trash box revealed that even the most simplistic robots can manage to evoke helping behavior in humans [170]. For HRI in public spaces, a first understanding of user acceptance related to robots as social actors was investigated in the Autonomous City Explorer (ACE) - project[1], where a robot navigates in urban environments, finding its way with the help of pointing gestures of human passers-by [164, 165], as a predecessor of the Interactive Urban Robot (IURO) - project[2], where also verbal interaction is explored in this context [28, 103].

[1] http://www.lsr.ei.tum.de/research/research-areas/robotics/ace-the-autonomous-city-explorer-project
[2] http://www.iuro-project.eu

In the research field of persuasive technology, Fogg [50] described five kinds of cues that have to be fulfilled to make humans think of robots as social actors:

1. *Physical cues*, like a face, eyes or a body
2. *Psychological cues*, like personal preferences, humor, personality, feelings, empathy, or the ability to apologize
3. *Language-related cues*, including the use of language and recognition of the same
4. *Social dynamics*, like taking turns in a game, cooperation, praising the user for good work, answering questions, and reciprocity, describing the concept of receiving and paying back favors
5. *Social roles*, with the robot being seen as a doctor, teacher or in a similar role

In order to achieve the goal of incorporating all social cues in interactive robots, there are extensive research efforts on building robot heads or robots with a full body that are able to express emotions: WE-3R III [148], WE-4R [172], Kismet [26], Leonardo [151], Roberta [65], Sparky [129], Feelix [29], Saya [81], Flobi [67], iCat [154], eMuu [7], Doldori [90], Ifbot [133], and Lino [153]. An increasing research interest is also observed regarding back-projected robot heads. This technique allows for smoother and subtler motions by means of projecting a 2D-screen avatar onto the back of a 3D-mask, e.g., LightHead [42], Furhat [105], and Mask-Bot [88]. A further trend in robotics research is to create lifelike "copies" of humans, so called androids, such as the Geminoid HI-1 [111], Geminoid-F [15] and the Geminoid-DK [157]. As related work shows, there are many different designs for emotionally expressive robots, whereas a differentiated validation of which is often pending. However, the legibility of the behavior of a robot is important for human users in order to interpret its intentions, and for the social cues to take effect. Accordingly, a systematical assessment of the design-dependent legibility of emotional expressions in speech and mimicry is conducted in this thesis in form of a comparative online-survey between the two different robot heads EDDIE and IURO in order to investigate the potential influence of machinelike versus humanlike design.

After validating the legibility of the IURO-head, an integrated dialog-architecture is implemented, where a social sub-dialog in terms of small-talk precedes the route inquiry of the robot. According to the experimental evaluation results in Chapter 3, explicit emotional adaption during small-talk, prior to task-related interaction, increases the prosocial behavior of human users towards a robot, as well as the perception on the HRI-key concepts of anthropomorphism and animacy.

Small talk, also called phatic communication, is described by Bickmore & Cassell as a talk where only interpersonal goals are discussed, while task goals are only marginally important [17]. As pointed out in [31] and [96], small talk opens up a way for humans to explore the ideas and beliefs of the conversation partner, thereby establishing the common ground, as discussed in Chapter 3, before entering the task-related part of the conversation. Phatic communication is deeply rooted in human rituals - though this kind of casual conversation is often led between strangers, e.g., a sales agent and a potential customer, it is used between close friends for the same reasons. The effectiveness is not restricted to finding a common ground: By enhancing the other person's social respect through appreciation, a sense of solidarity and thereby trust is built [17] to later carry the conversation to a deeper

level. It is also used as a means for maintaining social identities and relations [43]. In human-computer interaction (HCI), small talk is implemented in virtual agents - embodied computer programs with a human body or face as virtual display - to establish a prosocial common ground with users. A prominent example is REA, a conversational agent designed by the MIT as a real estate agent [17]. REA interviews potential buyers for properties, and tries to build up trust with the user for that goal. Small-talk plays an important role in this context, as results show that when revealing information about itself in small talk, REA is rated to be much more attractive by users, and users tend to rather buy a product from a small-talking agent. Bickmore also states the importance of relational agents - virtual agents designed to form a lasting relation with the user - but does not go into more detail. Stocky et al. [143] use small talk for grounding with MACK, another virtual agent. MACK gives route descriptions to passers by from his position in a kiosk, explaining a nearby paper map. Again, results point to much more lively conversations through the use of small-talk, and reduced attention by the user without it. Also cultural differences in small-talk are encountered in an application, where small talk is used to bridge gaps and awkward silences in international meetings by conversations between virtual agents and humans [71]. Up to now, not much research is conducted for small-talk in HRI. Many robots use small talk, but do not evaluate its specific influence on the interaction: Grace and George, two robots used as receptionist and guide at a conference [162], the seal robots used in elderly care [160] and Breazeal et al [24] use small talk for their robotic weight loss coach as a means for bonding and evaluation. Bartneck et al. [8] state phatic communication to be an important factor when judging the social abilities of a robot. This is consistent with Lee et al. [91], which evaluated human expectations when talking with a robot. Results show that people not only tend to greet human-like robots, but also use small talk with them during interaction instead of treating them like a non-social ticket automaton - even with no background knowledge about the robots abilities.

These implications go in line with the findings from Experiment V in Chapter 3. Implicit emotional adaption in form of an adaption in the emotional expressions of mimicry and voice increases empathy towards a robot. However, the explicit emotional adaption in an adaptive small-talk resulted in a significant impact on humanlikeness and helpfulness towards a robot. As Stocky et al. [143] stated, humans expect a robot or virtual agent with even roughly human looks to sport the abilities used in human communication. Even though the design of MACK, their virtual agent, is only partly humanlike (featuring a humanlike body, but a very robotic face), people show much more attention towards the agent when it displays its social capabilities by means of small talk. In the same way, the REA agent [17], which is more humanlike in its design, uses storytelling - a metaphoric form of phatic discourse - to reveal facts about herself, leading to much higher trust and raises efficiency for sales. Both projects share the principle behind their social strategy: They create a common ground with the user, which later is beneficial to the agent in consequent tasks and conversations. In this work, the same rule is transferred to IURO: While implicit emotional adaption is used in the experiments to increase its social capabilities and trigger empathy towards the robot, the common ground is created through a shared emotion during some the small-talk. This leads to heightened humanlikeness - as expected by Stocky et al. - and increased helpfulness, as the instinct to help is created through similarity and empathy with a social robot, as investigated in Chapter 3.

In order to ensure that the insights gained in Chapter 3, as investigated with the EDDIE-head, are transferable to the IURO-platform, a comparative study is conducted to inves-

tigate design-dependent differences in the legibility of emotional expressions between the two robotic heads EDDIE and IURO, as described in the following section.

4.2 Legibility of Emotional Robotic Expressions

Before being applied to the IURO-platform in combination with information retrieval in an outdoor field trial, the legibility of the emotions, expressed by IURO has to be evaluated. In other words, in order to transfer the findings of emotional adaption from an application on the EDDIE-head to the application on the IURO-robot, the legibility of the emotional expressions, generated in the same way, should at least not be worse than using the EDDIE-head. Thus, in order to ensure the legible transfer of emotional expressions animation, a comparative online survey is conducted prior to the outdoor field trials between the legibility of emotions shown by the two robotic heads EDDIE and IURO. Additionally, the user-perception of both robotic heads is evaluated towards the key-concepts of anthropomorphism and animacy without any interaction. Further, as a generalizable research question, a hypothesized impact of dispositional empathy on the human emotion recognition performance, as argued in [125], in identifying the depicted robotic emotions, is investigated in this section.

4.2.1 Experiment VI: Comparative Online-Survey on the Legibility of Emotional Robotic Expressions

A comparative video based online-survey is conducted in order to investigate the legibility of emotional speech in combination with the emotional facial expressions of two different robotic heads: EDDIE and IURO, animated identically, but using a different head-design as described in the following. For the evaluation, the emotions most relevant to the later information retrieval in the outdoor field trials are identified as happy, sad, surprised, and neutral as a control condition. The goal of the survey is to investigate three different topics:

1. Potential differences in the legibility of emotions as a function of head-design between EDDIE and IURO

2. Potential impacts of differences in dispositional empathy on the human performance in identifying the depicted emotions

3. Potential differences in the user-perception on the HRI-key concepts of anthropomorphism and animacy as a function of head-design between EDDIE and IURO

In the online-survey, the newly developed robotic head for the IURO-project is compared with the EDDIE-head. The IURO-head is based on the EDDIE-head and uses the same basic functionalities as described in Chapter 3, but a cover is added to make the robot robust against environmental outdoor impacts. In contrast to EDDIE, the IURO head-cover is more humanlike in its design, see Figure 4.2. The basic functionalities of IURO are: eye balls 2 DoF, eyelids 2x1 DoF, ears 2 DoF, mouth/jaw 1 DoF, lips 2x2 DoF. The head is built around commercially available miniature servo-mechanisms and the neck consists of 3 DC-motors equipped with harmonic drive gears.

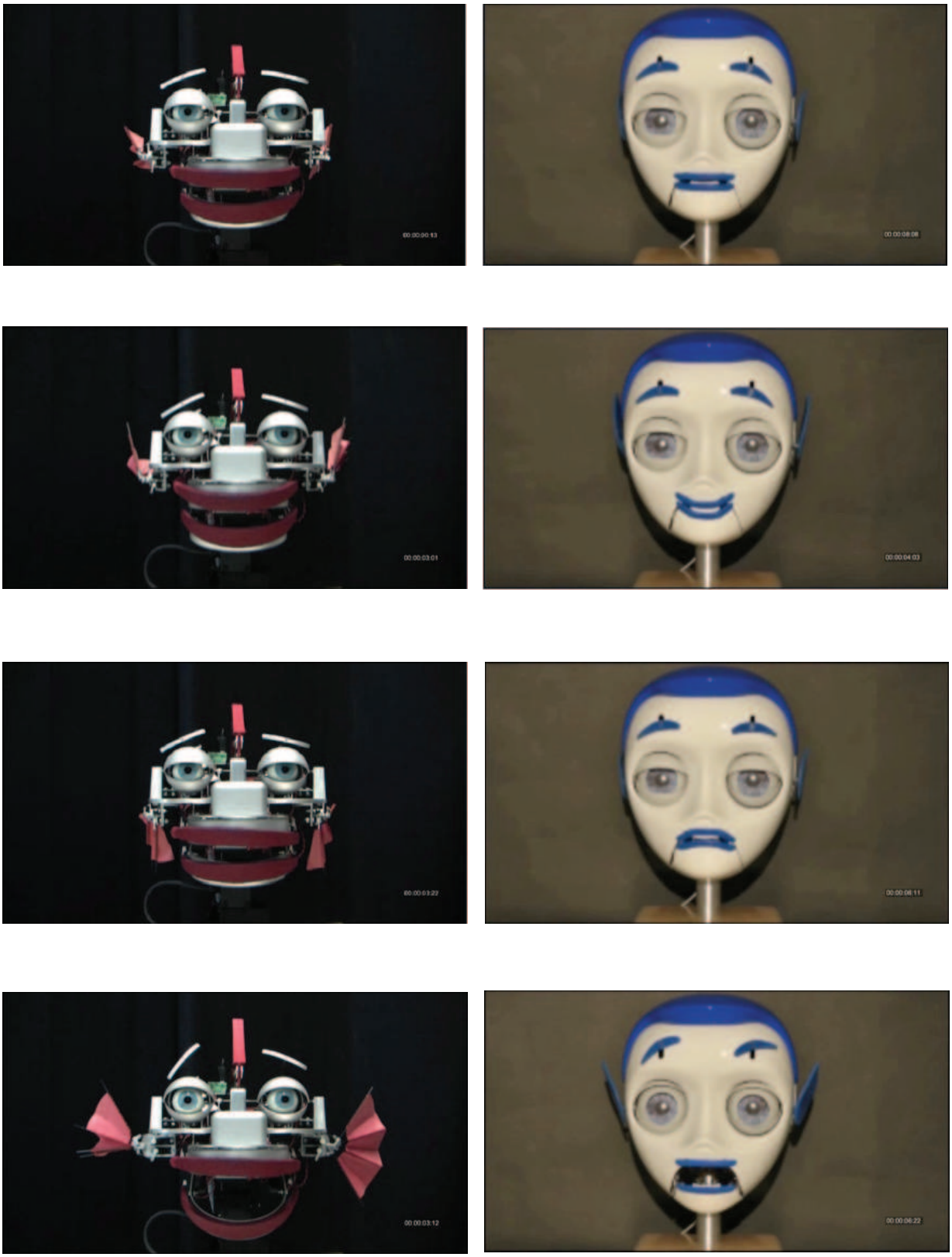

Figure 4.2: Experimentally evaluated emotions of EDDIE (left) versus IURO (right): neutral, happiness, sadness, surprise (from top to bottom)

For both robotic heads, the underlying model for the generation of facial expressions and emotional speech are based on the Pleasure-Arousal-Dominance (PAD) model [101], as already described more detailed in Chapter 3. In short, the PAD model represents emotions in a continuous three-dimensional space that is mapped to the joint space of the robot by converting the PAD values of the emotions to activations of facial action units (muscle groups) for emotional expressions, generated according to the Facial Action Coding System (FACS) [168] that allows an objective transfer of facial movements for the corresponding expressions to the robot face. Both, EDDIE and IURO use a subset of FACS including 13 action units relevant for emotional expressions. The underlying PAD values for the selected emotions are specified in Table 4.1

Table 4.1: Underlying PAD values for the selected emotions in the video-based online-survey from 1=very low to 5=very high

Emotions	P	A	D
Neutral	3.0	3.0	3.0
Happiness	4.5	4.0	3.5
Sadness	1.5	1.5	1.5
Surprise	3.0	4.5	2.5

The experimental design and used measures are described in the following.

4.2.2 Experimental Design & Measures

The video based online-survey is implemented and distributed over LimeSurvey[3], an open source survey software. After following the link to the survey, the subjects can choose to participate in the study either in German or English language. The survey consists of three different parts:

1. Personal data and self-assessment of dispositional empathy
2. Assessment of 8 randomized video-files of EDDIE and IURO
3. Concluding comparative assessment of anthropomorphism and animacy

The first questions to answer for the subject are of demographic nature, e.g., age and gender. After that, the subjects have to fill in the Toronto empathy questionnaire (TEQ), a self-assessment of their dispositional empathy [142], as already used in Experiment V in Chapter 3.

Subsequently, 8 video files showing the 4 selected emotions on EDDIE versus IURO of 10 seconds length are presented to the subject in randomized order. In all videos, both robots speak the sentence “I am a machine-like entity with emotions”, synthesized in an identical PAD-based emotional way as the corresponding facial expressions, matching the presented emotion as described in Chapter 3, Section 3.3.5. The videos are repeatable, and after watching each video, the subjects are asked if the video was played without disturbances. In case of a positive answer, each video is firstly assessed by using the SAM-questionnaire

[3]http://www.limesurvey.org

as depicted in Figure 3.11 in Chapter 3. In contrast to Experiment V, in this experiment the SAM-questionnaire is not used for a self-assessment of the user-mood, but to assess the legibility of the different PAD components of an emotion, presented by the robotic heads in the video. Additionally, a qualitative measure is included in form of an open question about the mood of the robot after each video-file. Subsequently, all subjects are asked to choose the emotion from a predefined list of different emotions, best matching that of the presented video-file. The emotion-list contains the actually depicted emotions, but is extended by additional emotions, not included in the experiment:

- Happiness
- Anger
- Disgust
- Fear
- Contempt
- Sadness
- Surprise
- Neutral

In this way, a total of 8 video-files are presented to be evaluated by the user.

In order to get a concluding comparative assessment of the anthropomorphism and animacy for the designs of EDDIE versus IURO, the corresponding Godspeed questionnaires [10] are employed at the end of the survey - interlinked with two additional pictures of the robotic heads.

4.2.3 Experimental Results

Experimental results are deduced from 73 subjects: 43 female and 30 male, aged between 12 and 54 years with an average age of 29.2 (SD 7.6), and different vocational backgrounds. No significant interaction effects were found between age and any other measure used in the survey.

Dispositional Empathy (TEQ)

Table 4.2 shows the mean values with standard deviations of the subjects for dispositional empathy (TEQ). Just like in Experiment V (Chapter 3), the means are slightly lower than the ones presented by Spreng et al. [142], proposing between 43 and 45 scores for male and between 44 and 49 for female participants, respectively. Thus, the male subjects are considerably below this prediction with a mean of 40,83 scores (SD 6.24). Accordingly, significant gender-differences are found: The means of the female subjects are significantly higher than the male means ($T = 3.898$, $p = .025$), but still just about their lower predicted boundary value with 44,52 mean scores (SD 5.23). As a consequence, the means of all subjects together are located at the lower boundary value for male subjects, although a higher amount of significantly more emphatic female subjects joined the survey in contrast to the previously conducted Experiment V. Nevertheless, an interaction effect was found

between dispositional empathy and the human performance in identifying the correct emotion when answering the qualitative measure by an open entry about the possible mood of the robot after each presented file for **Happiness** ($F = 4.69$, $p = .011$) and **Sadness** (($F = 3.54$, $p = .032$): The higher the dispositional empathy of the subjects the better is their identification performance of these emotions.

Table 4.2: Toronto Empathy Questionnaire (TEQ) mean scores (with a minimum of 0 and a maximum of 64 scores) and standard deviations (in brackets)

Subjects	TEQ Value
Female subjects	44.52 (5.23)
Male subjects	40.83 (6.24)
All subjects	43.00 (5.91)

SAM-Questionnaire

For the SAM-questionnaire, evaluating the legibility of the single PAD dimensions in the presented emotions, a repeated measures multivariate one-way ANOVA resulted in a significant influence of the factor head design ($F = 3.63$, $p = .001$) on the legibility of the PAD values (dependent variables) for the emotions. No interaction effects were found for dispositional empathy, age and gender. In the following, the detailed results are outlined for the presented emotions.

For **Neutral**, the mean values and standard deviations are depicted in Table 4.3.

Table 4.3: Neutral mean scores and standard deviations (in brackets) of the SAM-questionnaire on a 5-point Likert scale, compared to the original PAD values underlying the emotion generation

Condition	P	A	D
Original	3.0	3.0	3.0
IURO	2.75 (0.58)	2.07 (1.15)	2.42 (0.88)
EDDIE	2.94 (0.53)	1.76 (1.09)	2,65 (0.94)
n	72	72	72

As paired T-tests did not reveal any significant differences between EDDIE and IURO, as well as between both heads and the original PAD values for all dimensions of PAD, it can be deduced that the **Neutral** values of all PADs can be generally identified by the human subjects at both robotic heads.

For **Happiness**, the mean values and standard deviations are depicted in Table 4.4. Paired T-tests revealed significant differences between both heads, as well as between the both heads and the original PAD values in the dimension of *Pleasure* ($T \geq 4.500$, $p = .000$) and *Arousal* ($T \geq 4.000$, $p = .000$), with a trend of IURO being perceived closer to the original PAD values than EDDIE. For the dimension of *Dominance*, no significant differences were found between IURO and EDDIE. However, both heads are rated significantly different from the original PAD values ($T \geq 4.50$, $p = .000$), hinting to the fact that *Dominance* is rather difficult to identify in the implemented approach for emotional speech and facial expressions of both robotic heads.

Table 4.4: Happiness mean scores and standard deviations (in brackets) of the SAM-questionnaire on a 5-point Likert scale, compared to the original PAD values underlying the emotion generation

Condition	P	A	D
Original	4.5	4.0	3.5
IURO	4.04 (0.80)	2.78 (0.92)	2.88 (0.92)
EDDIE	3.39 (0.81)	2.35 (0.97)	2.71 (0.90)
n	72	72	72

For **Sadness**, the mean values and standard deviations are depicted in Table 4.5. Since the subjects had the possibility to skip a question in case of being not able to answer for any reason, the number of subjects (n) differs for this emotion for IURO, as indicated in the table.

Table 4.5: Sadness mean scores and standard deviations (in brackets) of the SAM-questionnaire on a 5-point Likert scale, compared to the original PAD values underlying the emotion generation

Condition	P	A	D
Original	1.5	1.5	1.5
IURO	1.90 (0.56)	2.2 (1.06)	2.37 (0.84)
n	73	64	73
EDDIE	1.93 (0.79)	2.13 (0.99)	2,39 (1.08)
n	72	72	72

Paired T-tests did not reveal any significant differences between EDDIE and IURO. However, a significant difference between both heads and the original PAD values are found for all PAD dimensions of *Pleasure*, *Arousal*, and *Dominance* ($T \geq 4.50$, $p = .000$), indicating that all PAD dimensions are hard to identify in the design of both robotic heads for **Sadness**.

For **Surprise**, the mean values and standard deviations are depicted in Table 4.6.

Table 4.6: Surprise mean scores and standard deviations (in brackets) of the SAM-questionnaire on a 5-point Likert scale, compared to the original PAD values underlying the emotion generation

Condition	P	A	D
Original	3.0	4.5	2.5
IURO	2.71 (1.13)	4.17 (0.93)	3.42 (1.16)
EDDIE	4.15 (1.00)	3.83 (1.09)	3.58 (1.00)
n	72	72	72

For *Pleasure*, paired T-tests resulted in significant differences between EDDIE and IURO, and between EDDIE and the original P-value ($T \geq 2.00$, $p = .000$). Since three T-tests are conducted, the significance level has to be adjusted from α=.05 to α=.017. As

a consequence, no significant difference exists between IURO and the original P-value for **Surprise** with $T \geq 2.00$, $p = .032$. In other words, the legibility of IURO was significantly closer to the original P-value while presenting the **Surprised** emotion. In contrast, the P-value of EDDIE was overestimated. For the dimension of *Arousal*, no significant difference was detected between EDDIE and IURO with $p = .024$, due to the adjusted significance level (α=.017), but between both robotic heads and the original A-value ($T \geq 3.00$, $p = .000$) with a tight trend towards IURO being closer to the original A-value of **Surprise**. In the *Dominance*-dimension, both robotic heads are rated significantly above the original D-value ($T \geq 6.50$, $p = .000$) by the subjects.

Further, it is analyzed if some PAD-dimension(s) was/were generally better identified than others when being presented by EDDIE versus IURO, independent from the depicted emotions. Therefore, an absolute error was calculated for each PAD-dimension by computing the absolute difference (above or below) between the original PAD values and those, identified in the presentations of EDDIE and IURO, respectively. As a result, paired T-tests revealed a significant difference of the absolute error between EDDIE and IURO for the dimension of *Pleasure* with $T = 5.70$, $p = .000$. The corresponding mean values of the absolute error in the dimension of *Pleasure* are 0.87 (SD 0.32) for EDDIE, and 0.65 (SD 0.28) for IURO.

As a summary of the PAD-analysis, it can be stated that the legibility of *Pleasure* is better for the head design of IURO than for the head design of EDDIE, independent from the depicted emotion. This may be due to the main differences in the emotions **Happy**, where a trend towards IURO being rightly rated as more pleasant than EDDIE was apparent, and **Surprise**, where EDDIE is erroneously rated more pleasant than IURO, resulting in a significant difference between EDDIE and the original PAD values. This is also visible in an overview of the PAD-ratings of EDDIE vs. IURO, compared with the original PAD values in Figure 4.3 and 4.4 . No interaction effect was found for age, gender, or dispositional empathy on the PAD-ratings in the SAM-questionnaire.

Open Guess & Predefined List of Emotions

After rating the single PAD dimensions in the SAM-questionnaire, that did not reveal the presented emotion itself, the subjects were firstly asked to make an open guess on how the robot is feeling in the video in form of an open entry. Secondly, a window with a predefined list of emotions appeared, where the subjects were asked to choose the emotion, best fitting to the presented video file. Both, the open entries and the selected emotions from the predefined list, were coded in either 1=correct or 0=wrong for each emotion and robotic head. The resulting human recognition rates are depicted in Table 4.7 and 4.8

Table 4.7: Human recognition rates of the emotions [%] by an **Open Guess**

	IURO	EDDIE
Neutral	64	60
Happiness	**67**	47
Sadness	70	78
Surprise	**69**	51
Mean recognition rate	**67.5**	59.0

For the open entries, paired T-tests revealed statistically significant differences in the human recognition performance between EDDIE and IURO for **Happiness** ($T = 3.01$,

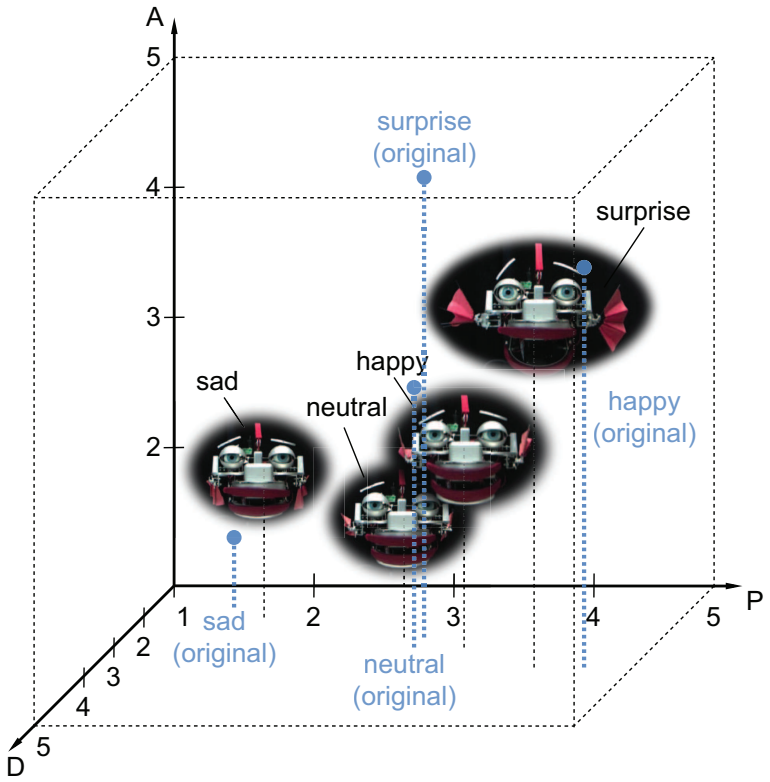

Figure 4.3: Overview of the PAD-ratings for EDDIE

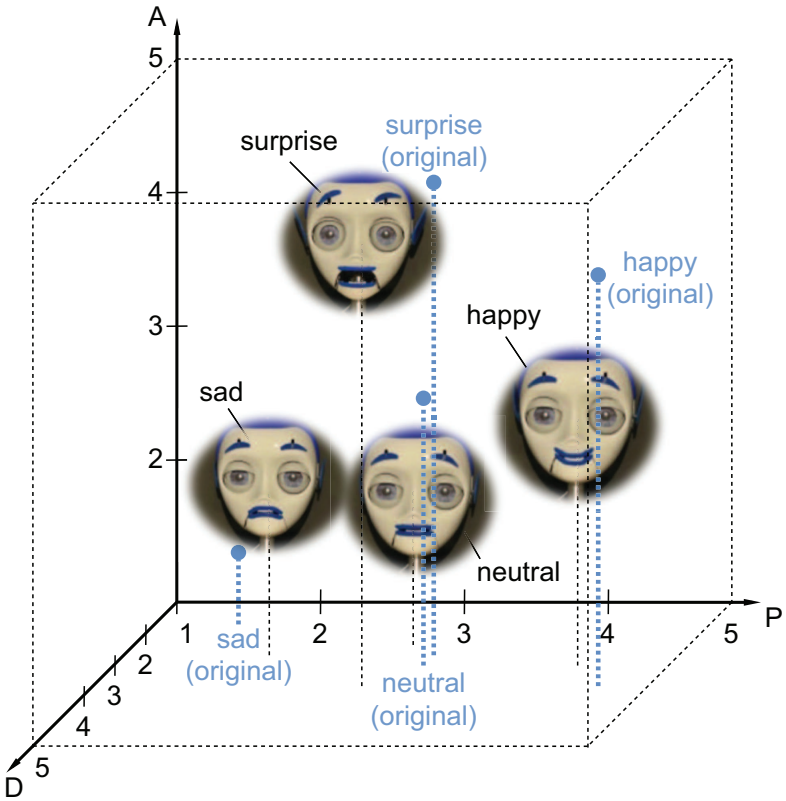

Figure 4.4: Overview of the PAD-rating for IURO

$p = .004$), and for the mean recognition rate in total ($T = 2.84$, $p = .006$). Although not statistically significant, a considerable difference is also notable for **Surprise** to the benefit of IURO, which supports the findings of the PAD-analysis, as described above. Further, a significant interaction effect was found for **dispositional empathy (TEQ)** on the human recognition rate of the emotions **Happiness** ($F = 4.69$, $p = .011$) and **Sadness** ($F = 3.54$, $p = .032$).

Table 4.8: Human recognition rates of the emotions [%] by a **Predefined Emotion-List**

	IURO	EDDIE
Neutral	76	83
Happiness	**65**	36
Sadness	68	74
Surprised	**53**	38
Mean recognition rate	65.5	57.8

The recognition rates of the predefined list of emotions go in line with those of the open entries, collectively. However, no statistically significant differences were found between the heads. This may be due to a higher distribution of individual differences caused by the high variety of emotions, offered in the list which may have triggered more room for interpretation in the subjects in contrast to the open entries.

Anthropomorphism & Animacy

Finally, an overall rating of EDDIE versus IURO was requested in the survey for the HRI key-concepts of anthropomorphism and animacy. Anthropomorphism was rated with a mean of 2.35 (SD 0.86) for EDDIE, and IURO could achieve a mean of 2.28 (SD 0.99) on a semantic differential scale from 1=very low to 5=very high. For animacy, the mean-ratings are 2.70 (SD 0.82) for EDDIE, and 2.47 (SD 0.90) for IURO. No significant head-design-differences were found, as well as no interaction effects.

4.2.4 Discussion

At the first glance, the ratings for anthropomorphism and animacy seem to be very low compared to the previous interactive experiments IV and V in Chapter 3. However, in the non-adaptive comparison group, where EDDIE did not employ either implicit or explicit emotional adaption, the mean value was comparably low with 2.36 (SD 0.69). By implicit emotional adaption, this mean could be increased to 2.73 (SD 0.76), by explicit emotional adaption to 3.07 (SD 0.72), and by applying both ways of emotional adaption the mean for anthropomorphism could be raised up to 3.13 (SD 0.76). Since no emotional adaption could be applied in this non-interactive online-survey, the low means are plausible. For animacy, the means of the previous experiments were higher in all tested conditions, ranging from 3.10 (SD 0.7) in the neutral condition of Experiment IV up to 3.82 (0.58) in Experiment V, where both, implicit and explicit emotional adaption, were applied. The lower ratings in the presented online-survey for EDDIE with 2.70 (SD 0.82), and IURO with an even lower mean of 2.47 (SD 0.90) may be due to the missing interactive part of the experiment: Since animacy is very much bound to interactive animation and reactivity, this may not be conveyed through the presented videos, which could only be watched in a passive way by the subjects, and no interaction effects could appear in this setting.

The PAD-analysis of the emotions resulted in a significantly better legibility of IURO on the dimension of *Pleasure*, which particularly affects the emotions **Happiness**, where a trend towards IURO being rated correctly as more pleasant than EDDIE was noticeable, and **Surprise**, where EDDIE is erroneously rated more pleasant than IURO, resulting in a significant difference between EDDIE and the original PAD values. In contrast, the dimension of *Dominance* was rather difficult to identify in the facial expressions and prosody in speech of both heads. Interestingly, this is a vice-versa relation to the results, achieved by Karg [77], where an investigation of the legibility of emotional gait patterns revealed a good legibility of *Dominance*, and a rather poor legibility of *Pleasure*. Thus, it can be deduced that *Dominance* may be better expressed by body movements than by facial mimicry and/or prosody in speech. Correspondingly, the legibility of *Pleasure* seems to be highly bound to mimicry and speech-prosody.

Independent from the single PAD-dimensions, for the open guess on how the robot may feel, and the predefined emotion-list where the subjects had to choose the matching emotion, the same effect was visible: The recognition rates of **Happiness** and **Surprise** were up to 20% better for IURO. The fact that the recognition rates as a whole are lower than in the pretest of Experiment V with EDDIE, depicted in Table 3.8, may be due to the fact that in the pretest only 20 staff members of Technische Universität München (TUM) were chosen as subjects. Hence, the sample was potentially used to robots and, thus, showed a better recognition performance. In this sample, however, 73 subjects with very different backgrounds participated in the survey, thereby better representing potential users. However, just like for anthropomorphism and animacy, another aspect is the missing interactive part in the presented online-survey. It is presumed to be easier for subjects to interpret emotional speech and facial expressions in an interactive situational context. Thus, it is expected that the recognition rates of human users will increase again in a real interaction scenario, where a task-related interaction context provides an added value for the identification of emotions, as given in the outdoor field trials described in Section 4.3.

An interaction effect of dispositional empathy on the human recognition performance was only found for the open entries, where the subjects had to guess how the robot may feel in the video. Hence, dispositional empathy only affects the recognition performance in open guesses for **Happiness** and **Sadness**, which are located opposed to each other in the dimension of *Pleasure*. Thus, it may be deduced that dispositional empathy affects the human emotion-recognition performance in the dimension of *Pleasure*. However, the influence is only given when the interpretation is not supported, and thus not triggered, by any preset measure, e.g., by pictures in the SAM-questionnaire or a predefined emotion list to choose from, but rather if the recognition of an emotion is solely based on an intuitive sense that benefits from dispositional empathy in case of much room for interpretation. Accordingly, no significant influence of dispositional empathy on recognition performance is given or the predefined emotion list, maybe due to the constrained interpretation room by the variety of preset emotions to choose from.

After validating the legibility of the IURO-head, an integrated dialog-architecture is developed and implemented in the IURO-platform, where a social sub-dialog in terms of some small-talk precedes the route inquiry of the robot. The application of the fully integrated approach in outdoor environments is presented in the following section.

4.3 Application of the fully Integrated Approach in Outdoor Environments

In this section, a fully integrated approach is developed and implemented in the robotic system IURO. For the integrated approach, proactive information retrieval and prosocial behavior control are combined. The emotional adaption approach is extended by an emotion recognition module to estimate the user-mood before adapting to the same. In this way, emotional adaption is integrated in the robotic system to enable the robot to align with humans in a fully automated way during information retrieval. Thus, the approaches for proactive information retrieval and emotional adaption, as presented in the previous chapters, are combined in the dialog system to allow for prosocial information retrieval in outdoor environments. An outdoor field trial is conducted in order to evaluate the integrated approach. The experimental results reveal a significant gender effect in user experience for subjective system-performance and perceived enjoyment: A trend of higher ratings of female users emerged for perceived enjoyment, whereas male users came up with higher ratings on subjective system-performance. Further, a significant interaction effect is found between emotion recognition and speech recognition quality: Since the willingness of human users to help the robot only significantly decreases if both, emotion and speech recognition modules showed poor performance, it can be deduced that the two modules are compensating each other towards the induction of helpfulness, as long as recognition works for at least one of the two modules.

4.3.1 Integrated Architecture

A main challenge, addressed in this work is to motivate unconcerned passers-by to help a robot with missing route-knowledge for a task they do not benefit from. Thus, for a robot, a more complex task may be divided into different cognitive and social sub-tasks that have to be achieved in order to fulfill its task, e.g., a social sub-task for a robot could be to actively create a prosocial situational context in order to motivate people that are unconcerned with its task to be cooperative and help out with missing task-knowledge. Since robots cannot be pre-programmed with all the world-knowledge humans possess, humans are an indispensable information source for robots operating in dynamic real-world scenarios. Thus, robots may need to accomplish a social sub-task first, in order to fulfill the cognitive sub-task to retrieve missing task-information before it can execute the task itself, see Figure 4.5.

In the integrated architecture, proactive information retrieval and prosocial behavior control are combined. According to the experimental results in the previous Chapter 3, explicit emotional adaption in form of an adaptive social sub-dialog is most powerful for prosocial behavior control. Thus, the approach is extended by an emotion recognition module in order to expand the effect by guessing on the mood of a human user, before a similarity statement is uttered to adapt to the same. In order to create a prosocial basis for information retrieval, the social sub-dialog is integrated in the dialog system, precedent to obtaining missing task information from humans.

Emotional adaption is implemented in form of a social sub-dialog (SSD) prior to the route dialog. The social sub-dialog itself incorporates explicit adaption towards the mood of the user: An inquiry about the user-mood is followed by a similarity statement of

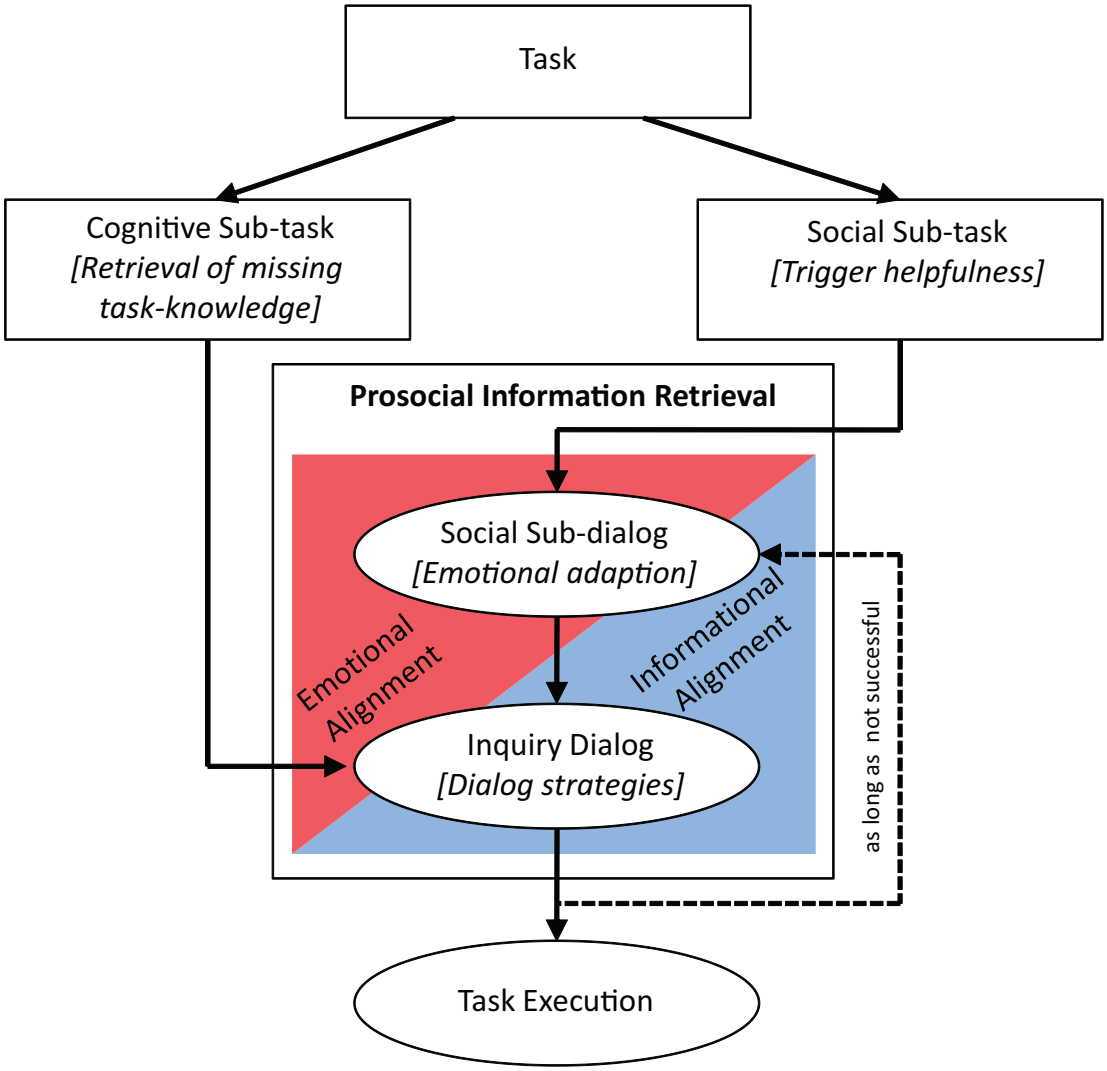

Figure 4.5: General framework for task-related HRI with missing task-knowledge, split up into a cognitive and a social sub-task

the robot being in the same mood. At the same time, the emotional state of the robot, underlying the synthesis of emotional mimicry, is shifted by a bias towards the mood of the user for the following route dialog. The implementation uses XML-files that define a Finite State Machine (FSM) incorporating the dialog structure, as depicted in Figure 4.6:

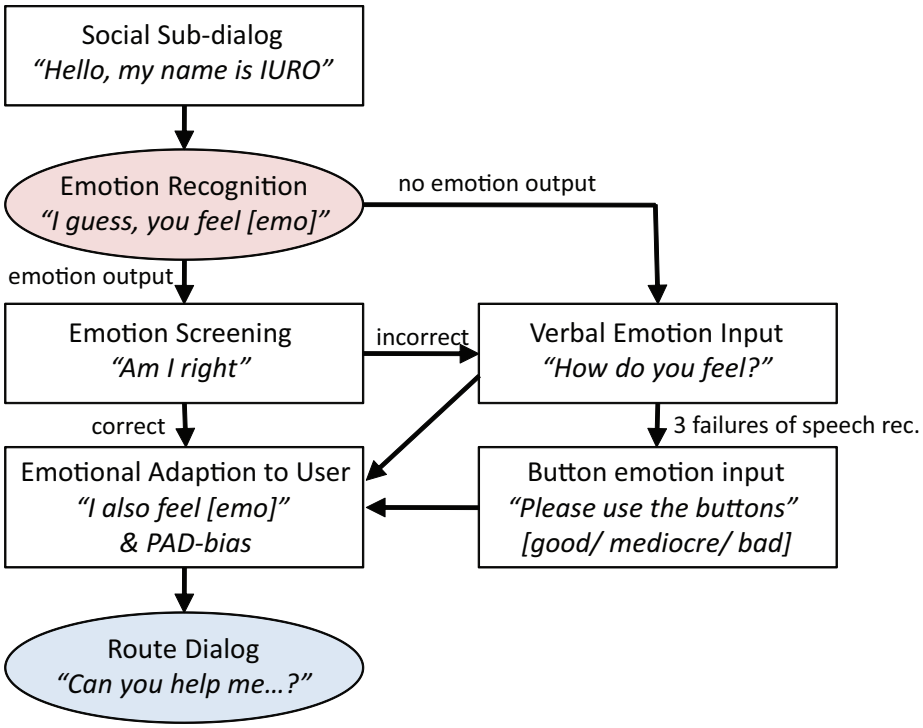

Figure 4.6: Integrated dialog structure of emotional adaption in form of a social sub-dialog

In order to apply emotional adaption, the original FSM of the IURO-dialog-system is enhanced by additional states. The original initial state controls the detection of users and starts all associated states as well as returning the state in case of unexpected dialog abortion. This initial state is changed to the requirements of emotional adaption in terms of a variable initiating the social sub-dialog to increase helpfulness towards the robot. Thereby, a distinction is made between cases where the emotion of the user could be determined based on a facial expressions-analysis by the emotion recognition module Shore [4], and cases where emotion recognition did not work initially. In case of an existing emotion recognition output the FSM proceeds in a state, where IURO expresses its estimation of the current mood to the user, and screens the emotion of the user accordingly. Shore uses values between 1 and 100 for the basic emotions happy, angry, neutral, sad, and surprised, that can be interpreted as confidence values. The emotion with the highest confidence is used for an estimation of the user-mood in the dialog system. The FSM screens the estimated emotion by the inquiry: "I guess you currently feel [emo]?". In case of a successful guess by the robot, the following emotional expressions of the robot are adapted to this emotion in terms of being shifted by a corresponding bias in the underlying emotional representation on the PAD dimensions [101]. In case of an incorrect estimation of the user-mood, the user is asked for a verbal emotion input with the utterance "how do you feel then?", identically with cases, where no emotion could be detected in the facial expressions of the human

[4]http://www.iis.fraunhofer.de/de/bf/bsy/produkte/shore.html

interaction partner. In case of three failures caused by bad speech recognition quality, the emotion can be inserted via predefined buttons on a touch screen.

In any case, the robot answers with the similarity statement "me too", and shifts its baseline for further emotional expressions towards the corresponding PAD values, transmitted to the mimicry of the robot via ROS. In this way, the robot is able to sustain implicit emotional adaption for the interaction.

A simulation environment is developed, which facilitates the simulation of messages, so that changes of the dialog structure and exchange of variables between the modules of the robot can be tested. An event-based exchange of data is realized by the FSM-structure, which controls speech recognition, robot motion, and speech output. For the social sub-dialog, a modification of the variable structure is necessary in order to add emotional information in form of a new event containing this emotional information. This event sets the current emotion to be displayed by the facial expression system through the FSM. Additional six event structures are added, covering recognized emotions from the visual emotion recognition module of the robot.

The integrated architecture is applied and evaluated in a fully integrated system in an evaluative outdoor field trial, as described in the following.

4.3.2 Experiment VII: Prosocial Information Retrieval from Humans in an Outdoor Environment

In order to evaluate the feasibility of the integrated approach, an outdoor field trial is conducted with the robotic platform IURO, where the emotional adaption approach to trigger helpfulness in humans towards a robot is paired with retrieval of missing task-knowledge by natural language HRI. The experimental design and measures are described in the following.

4.3.3 Experimental Design & Measures

The fully integrated outdoor experiment was distributed over six runs, which were conducted in October 2012. Each run lasted approximately 3-4 hours depending on the battery discharging. As can be seen in Figure 4.7, the experimental design of the outdoor field trial can be described with the robotic platform IURO driving around in the city center of Munich, proactively approaching passers-by and initiating natural language interactions with approximately 100 people by firstly making a guess about their mood, and secondly asking for the way to a public place. In 36 cases, the interaction led to a full route dialog.

In order to increase the willingness to help, the route dialogs are opened by small talk about the current mood in a social sub-dialog, where emotional adaption is applied before asking for directions. All interaction partners are observed plus video-taped, and interviewed with questionnaires on user experience (UX), subjective system performance, and social acceptance.

Questionnaires:

In order to evaluate the fully integrated architecture with emotional adaption implemented in the social sub-dialog and combined with asking for directions, all interaction partners (30 out of 36) that had a full social sub-dialog followed by a full route dialog with IURO, filled in these questionnaires. The questionnaires consist of a selected combination of different

Figure 4.7: Impressions of the robotic platform IURO in the outdoor field trial

state-of-the-art and newly developed measure-constructs on user acceptance, willingness to help, subjective task-performance, anthropomorphism, and animacy of the robot as well as the degree of situational empathy induced by the robot during the interaction:

- The previously developed (see Chapter 3), but adapted questionnaires on situational empathy and subjective performance
- A newly developed questionnaire-construct on the willingness of humans to help a robot
- An adapted selection of state-of-the-art constructs for user acceptance [66], i.e., perceived sociability, social presence, perceived enjoyment, and intention to use.
- The unmodified Godspeed-constructs for anthropomorphism and animacy [10].

The questionnaires (without the unmodified Godspeed-constructs) are depicted in Table 4.9.

Video Annotation:
A video analysis is conducted for two performance-dependent influence factors on the interactions:

- Emotion recognition quality (successful/ misrecognition/ non-recognition: IURO asked for user-mood)
- Speech recognition quality (successful/ miscommunication/ non-recognition: Input via buttons)

4.3.4 Experimental Results

Experimental results could be deduced from 30 subjects, that had both, a full social subdialog and a full route dialog with IURO. The mean age of these people was 36.9 years (SD 18.9), ranging from 13 to 74 years. 18 of the respondents were male, 12 female.

For the newly developed questionnaire-construct on the human willingness to help, Cronbachs α was calculated to evaluate the internal reliability of the items. As solid construct should create a Cronbachs $\alpha \geq 0.70$, the items of the novel construct showed an acceptable reliability with Cronbachs $\alpha = 0.73$.

	Situational Empathy
1)	I would be happy for IURO if it reaches its goal location.
2)	It would be a pity if IURO gets lost underway.
3)	It would be funny if IURO gets lost underway. (inverted)
4)	I would feel sorry for IURO if someone tried to destroy it at that moment, thus I would try to prevent it.
	Subjective Performance
1)	I had the feeling that IURO understood my directions.
2)	IURO has shown a good performance.
3)	I think that IURO has worked efficiently.
4)	It took IURO (too) long to understand my directions. (inverted)
	Willingness to Help
1)	Humans should help a robot to fulfill its task.
2)	Robots should not ask humans for help. (inverted)
3)	I really wanted to help IURO to find its goal location.
4)	I would help IURO again to find its goal location.
	Perceived Sociability
1)	I like IURO.
2)	IURO's mimic and verbal statements fit together well.
3)	IURO was a good conversation partner.
4)	IURO's behavior was inappropriate. (negated)
	Social Presence
1)	I had the feeling that IURO really looked at me.
2)	I could imagine IURO as a living being.
3)	Sometimes it felt like IURO had real feelings.
4)	IURO's behavior was not humanlike. (negated)
	Perceived Enjoyment
1)	It was fun to interact with IURO.
2)	The conversation with IURO was fascinating.
3)	I consider IURO to be entertaining.
4)	It's boring when IURO interacts with me.(negated)
	Intention to Use
1)	I would like to interact with IURO more often.
2)	I would take IURO home with me.
3)	I would like to play again with IURO within the next few days.
4)	I could imagine interacting with IURO over an extended period of time.

Table 4.9: Questionnaires for User Acceptance on a 5-point Likert scale, extended by a new construct on the willingness to help a robot

Table 4.10 shows the means and standard deviations (SD) of the user ratings for the constructs measuring all interactions with applied emotional adaption on a 1: very low 5: very high Likert scale.

Table 4.10: Means and standard deviations (SD) of the user ratings for the constructs measuring all interactions with applied emotional adaption (on a 1: very low 5: very high Likert scale)

Construct	**Mean (SD)**
Situational Empathy	**4.15** (0.89)
Subjective Performance	3.52 (0.95)
Willingness to Help	**4.09** (0.93)
Perceived Sociability	3.93 (0.78)
Social Presence	2.87 (0.94)
Perceived Enjoyment	4.18 (1.01)
Intention to Use	3.77 (1.25)
Anthropomorphism	2.92 (0.86)
Animacy	4.18 (0.75)

Since high situational empathy has to be assumed for emotional adaption to lead to increased helpfulness towards a robot [18], this assumption can be regarded as fulfilled with a mean of 4.15 (SD 0.89). Accordingly, the willingness to help construct resulted in a comparable high mean of 4.09 (SD 0.93). Correlation analysis focused on empathy and willingness to help along with the other constructs: For both, empathy and willingness to help, correlations were revealed to each other (p=0.002) and all other constructs ($p \leq 0.020$), except of social presence. Univariate two-way ANOVAs for all constructs using job (working - in education - not working/retired - technical background) and gender (male - female) as independent variables revealed a significant influence of gender on subjective performance (F=4.764, p=0,039) and perceived enjoyment (F=4.866, p=0.039). Post hoc T-tests between male and female users did not reveal significant group differences for subjective performance, but a nearly significant trend for perceived enjoyment (F=1.877, p=0.071) to be higher for female users (mean=4.61, SD=0.82) than for male users (mean=3.92, SD=1.05).

As described above, emotional adaption was opened by an estimation of the robot on the current user mood in the social sub-dialog. After the user either confirmed or declined this estimation via speech recognition or buttons (in case of bad speech recognition quality), IURO adapted its mood to the user by an explicit statement, and by shifting the baseline of the implicit generation of emotional facial and verbal expressions by a bias according to the user mood. Thus, in order to evaluate the effects of the performance shown during the social sub-dialog, two potential influence factors on the user ratings were identified as independent variables with three stages, deduced from observations and video analysis: Firstly, the emotion recognition quality regarding the user mood, including the stages of successful recognition: 22 cases, misrecognition: 4 cases, non-recognition: 4 cases (e.g. because of bad light conditions), where the users had to be asked about his/her mood. The second factor is speech recognition quality with the three stages of successful recognition: 14 cases (where IURO understood the user input immediately), miscommunication: 8 cases (where IURO needed an inquiry), non-recognition: 8 cases (when speech recognition did

not work at all and the users had to use buttons for their input). A univariate two-way ANOVA on the constructs revealed a significant influence for emotion recognition quality on willingness to help (F=5.812, p=0.009), as well as an interaction with the factor speech recognition quality (F=4.122. p=0.018) with regard to a decrease of Willingness to help in case of misrecognition of emotion paired with non-recognition of speech (buttons). For subjective performance, a univariate two-way ANOVA also revealed a significant influence of emotion recognition quality (F=4.407, p=0.025), but without any interaction effects to speech recognition quality.

4.3.5 Discussion

In summary, the results of the questionnaire of emotional adaption revealed acceptable reliability for the novel construct of willingness to help, resulting in high mean ratings for this construct and all other constructs. Correlation analysis revealed positive correlations between all constructs, except of social presence, that additionally resulted in the lowest mean value compared to the other constructs (just like in previous work [60]). The Job of the users did not influence their ratings of the interaction. The experimental results reveal a significant gender effect in user experience for subjective system-performance and perceived enjoyment: A trend of higher ratings of female users emerged for perceived enjoyment, whereas male users came up with higher ratings on subjective system-performance. Further, a significant interaction effect is found between emotion recognition and speech recognition quality: The willingness of human users to help the robot only decreases, if a false estimation (misrecognition) of the user-mood is paired with bad speech recognition quality (non-recognition). By implication, if only one of the modules results in mis- or non-recognition, the willingness to help does not increase. Thereby, the two modules of emotion recognition and speech recognition are compensating each other as long as the recognition performance works for at least one of the modules. In order to quantify the impact of this relation, in the outdoor field trials, speech recognition was successful in approximately half of the interactions, and emotion recognition worked out in circa in two-thirds of the interaction.

4.4 Summary

In this chapter, the combination of both, informational and emotional alignment, into an integrated approach for prosocial information retrieval in urban outdoor environments was applied on the robotic platform IURO in a field trial in outdoor environments.

Since emotional alignment is highly associated with the legibility of emotional expressions, as a first step, a comparative video-based online-survey was conducted to reveal potential design-dependent differences between EDDIE and IURO. The results indicate a better human recognition rate for IURO than for EDDIE, especially on the dimension of pleasure. Further, the survey revealed a significant interaction between dispositional empathy and open guesses of the subjects about the mood of the robot for the emotions happiness and sadness, again hinting to a key-function of the pleasure-dimension for the legibility of emotions. Further, an interesting insight was provided for anthropomorphism: The more humanlike design of the IURO-head does not necessarily result in a higher anthropomorphism-rating of the same. Thus, the online-survey reconfirms the importance

of the interactive behavior of a robot in a situational context, as well as the positive impact of emotional alignment as a consequence thereof.

Since the results of the online-survey indicated a benefit in the legibility of emotions for IURO, the use of the IURO-head is justified for the evaluation of the integrated approach in outdoor field trials. For the integrated architecture, as presented in this chapter, proactive information retrieval and prosocial behavior control were combined. The emotional adaption approach is extended by an emotion recognition module to estimate the user mood before adapting to the same. Thus, emotional adaption was integrated in a robotic system to enable the robot to align with humans in a fully automatic way during information retrieval. The experimental results of the outdoor field trial revealed a significant gender effect in user experience for subjective system-performance and perceived enjoyment: A trend of higher ratings of female users emerged for perceived enjoyment, whereas male users came up with higher ratings on subjective system-performance which is related to the willingness of human user to help a robot. Further, a significant interaction effect is found between emotion recognition and speech recognition quality: Since the willingness of human users to help a robot only decreases if both, emotion and speech recognition modules showed poor performance, it could be deduced that the two modules are compensating each other for helpfulness, as long as recognition works for at least one of the two modules. Summing all up, it can be deduced that proactive information retrieval in outdoor environments benefits from being combined with emotional adaption, since all tested dimensions of user experience (UX) resulted in comparably high mean ratings, which are positively related to the willingness to help a robot. Thus, the combination of informational and emotional alignment brings a benefit to task-related HRI by reinforcing the underlying prosocial motivation of humans to help a robot, thereby even compensating for bad speech recognition performance. Hence, the integration of the emotional alignment strategies as developed in this thesis in future social robots that are operating and retrieving information in outdoor environments is of high benefit for user acceptance and the human willingness to cooperate with those robots.

5 Conclusions and Future Directions

5.1 Concluding Remarks

The introduction of domestic robots into the real world faces a variety of interdisciplinary issues. In particular, user acceptance and the willingness of humans to cooperate and interact with robots have to be maintained. In turn, robots should be aware of their situational knowledge limitations and be able to pro-actively and flexibly acquire the knowledge needed to perform the objectives given by their human masters.

Thereby, a robot has to cope with various environmental impacts, e.g. noisy outdoor conditions. In order to overcome this bottleneck of speech recognition, different dialog strategies and specified miscommunication handling requests are developed and experimentally evaluated in this dissertation. In order to adapt to varying speech recognition performance while maintaining highest possible naturalness to the user, a switching mechanism is developed that allows smooth transitions between open and closed requests. The dialog strategies are embedded in a framework for pro-active information retrieval, also incorporating hypothesis-driven information processing, as well as the representation and evaluation of the acquired knowledge during task-execution within real world.

User acceptance and the willingness to cooperate and interact with a robot have been increased by a targeted integration of psychological interaction mechanisms, modeled according to theories from social psychology. In human-human interaction, empathy and a feeling of having something in common with a person in need of help, e.g. in personal attitudes, are essential motivational influence factors for prosocial behavior. Thus, different emotional control variables have been developed and integrated in a behavior control model. In evaluative experiments, emotional behavior control was successfully applied to proactively trigger these feelings in human users. In particular, this is achieved by pro-active small-talk mechanisms, employed prior to task-related interaction in order to establish a prosocial and cooperative common ground, generalizable for any human-robot interaction. Additionally, the social capabilities of the robot are enhanced by corresponding emotional facial expressions and prosody in speech throughout the interaction in a way, emotionally adaptive to the user.

Since emotional alignment is highly associated with the legibility of emotional expressions, a comparative video-based online-survey was conducted, before testing the generalizability of the approach in an integrated architecture for prosocial information retrieval with the IURO-platform. Thereby, a differentiated assessment of the design-dependent legibility of emotional expressions in speech and mimicry was conducted between two differently designed robotic heads of either machinelike versus more humanlike design. Also, potential differences in the user-perception on the HRI-key concepts of anthropomorphism and animacy have been considered. Additionally, impacts of dispositional empathy on the human performance in identifying the animated emotions are revealed, and the importance of an interactive context for emotion recognition is confirmed.

In order to integrate the investigated aspects of informational and emotional alignment, a general underlying framework has been developed integrating both, social and cognitive sub-tasks. Therein, informational alignment is integrated in the cognitive sub-task of information retrieval, interlinked with emotional alignment in the social sub-task of triggering helpfulness in human users in an implicit an explicit way of prosocial behavior control.

An integrated dialog architecture for prosocial information retrieval was developed and implemented in the dialog system of the robotic IURO-platform to be used in an evaluative outdoor application of the integrated approach. Thus, proactive information retrieval and prosocial behavior control are combined in form of a social sub-dialog prior to the route inquiry-dialog: While implicit emotional adaption in terms of emotional facial mimicry is used to increase the social capabilities and trigger empathy towards the robot, a common ground of the interaction is created by the integration of small-talk before entering the task-related part of the interaction. Moreover, prosocial behavior control is extended by an emotion recognition module in order to align with humans in a fully automated way during information retrieval.

The fully integrated approach of prosocial information retrieval is evaluated in an outdoor field trial. The experimental results confirm the positive effects of combining informational and emotional alignment with human users since all tested dimensions of user experience (UX) resulted in comparably high mean ratings, positively related with the willingness of humans to help a robot. A significant interaction effect between the technical performances of emotion- and speech recognition showed that the willingness to help a robot only decreases if both, emotion and speech recognition modules, show poor performance. Thus, it can be deduced that informational and emotional alignment are compensating each other with regard to helpfulness: In case of poor speech recognition performance in outdoor environments, successful emotional alignment keeps up the interest of humans to cooperate with a robot.

Summarizing, it can be deduced that proactive information retrieval benefits from being combined with emotional adaption, since it reinforces the underlying prosocial motivation of humans to help a robot, thereby even compensating decreases in user acceptance due to bad speech recognition performance. Accordingly, the ideas, concepts, and approaches developed in this thesis significantly advance the state of the art in design and control of social HRI and information extraction from natural language.

5.2 Outlook

The research field of social robotics is of highly interdisciplinary nature. By combining the research fields of linguistics and social psychology with computer science and robotics engineering, the work in this thesis provides a solid ground for future interdisciplinary research on information retrieval from human-robot interaction (HRI) and prosocial behavior control. The topics addressed in this dissertation also motivate a number of interesting future research directions, as drafted in the following.

- *Social-psychological interaction control models* - This work showed that theories from social psychology on human-human interaction are transferable to HRI. In this dissertation, a behavior control model was exemplarily defined for the social sub-task of helpfulness, needed in this context, but can be enhanced by additional social sub-tasks, specified for any other tasks.

- *Dynamic behavior control* - Each interaction can be seen as a dynamic process. Yet, dynamic emotional changes in the course of the interaction are not considered in the behavior control model. Thus, an interesting field of future research is the integration of system-theoretic approaches to prosocial behavior control in terms of applying dynamic mathematical models on emotional alignment, deduced from statistically gained interaction data.

- *Predictions on informational alignment in dialog strategies* - The switching mechanism, developed in this thesis, allows an adaption of the dialog strategy on varying speech recognition performance by monitoring the amount of miscommunication versus the amount of correctly extracted information. One future aspect is to enhance the approach by an integration of stochastic mathematical models in order to predict informational alignment over time, and the application of tools from control theory, allowing for not only reactive, but also preventive switching dialog strategies.

Many aspects of informational and emotional alignment of the research presented in this dissertation are not limited to natural language HRI, but are basically applicable to any task-related HRI, exploiting the mentioned benefits. Research on informational and emotional alignment strategies will have a large impact on integrated concepts of multi-modal interactive systems. Promising will be the joint research on haptic collaborative systems, where significant synergies are expected due to a high transferability and applicability of the developed communication strategies to haptic communication channels, which will highly advance the state of the art with high impact on future technology and applications.

Bibliography

[1] J. Ahlberg. Candide-3 – an updated parameterized face. Technical report, Linköping University, Sweden, 2001.

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

[3] S. Asteriadis, P. Tzouveli, K. Karpouzis, and S. Kollias. Estimation of behavioral user state based on eye gaze and head pose - application in an e-learning environment. *Multimedia Tools and Applications*, 41(3):469–493, 2009.

[4] R. W. Backs, J. K. Lenneman, J.M. Wetzel, and P. Green. Cardiac measures of driver workload during simulated driving with and without visual occlusion. *Human Factors*, 45(4):525–538, 2003.

[5] J. N. Bailenson and N. Yee. Digital chameleons: automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychol Sci*, 16(10):814–819, 2005.

[6] L. Barsalou, C. Breazeal, and L. Smith. Cognition as coordinated non-cognition. *Cognitive Processing*, 8:79–91, 2007.

[7] C. Bartneck. Interacting with an embodied emotional character. In *Proceedings of the International Conference on Designing Pleasurable Products and Interfaces*, pages 55–60. ACM Press, 2003.

[8] C. Bartneck and J. Forlizzi. A design-centred framework for social human-robot interaction. In *Proc. of Int. Workshop of Robot and Human Interactive Communication (RO-MAN)*, 2004.

[9] C. Bartneck, D. Kulic, and E. Croft. Measuring the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. In *Proc. Workshop on Metrics for Human-Robot Interaction*, pages 37– 44, Amsterdam, 2008.

[10] C. Bartneck, D. Kulic, and E. Croft. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1):71–81, 2009.

[11] C. D. Batson, B. D. Duncan, P. Ackermann, T. Buckley, and K. Birch. Is empathic emotion a source of altruistic motivation? *Journal of Personality and Social Psychology*, 40:290–302, 1981.

[12] A. Bauer, B. Gonsior, D. Wollherr, and M. Buss. Heuristic rules for human-robot interaction based on principles from linguistics - asking for directions. In *Proc. of AISB Convention - Int. Symp. on New Frontiers in Human-Robot Interaction (Best student poster award)*, pages 24–30, 2009.

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

[14] V. Bauwens and J. Fink. Will your household adopt your new robot? *Interactions*, 19(2):60–64, 2012.

[15] C. Becker-Asano and H. Ishiguro. Evaluating facial displays of emotion for the android robot geminoid F. In *Proceedings of the IEEE Workshop on Affective Computational Intelligence (WACI)*, pages 1–8, 2011.

[16] D.M. Berger. *Clinical empathy*. Jason Aronson, Inc., Northvale, 1987.

[17] T. W. Bickmore and J. Cassell. Small talk and conversational storytelling in embodied interface agents. In *Proc. of the AAAI Fall Symposium on Narrative Intelligence*, Cape Cod, MA., 1999.

[18] H.-W. Bierhoff. *Theorien der Sozialpsychologie: Gruppen, Interaktions- und Lerntheorien*, chapter Theorien hilfreichen Verhaltens, pages 178–197. Huber, Bern, 2002.

[19] S. J. Blakemore, J. Winston, and U. Frith. Social cognitive neuroscience: where are we heading? *Trends in Cognitive Sciences*, 8, 2004.

[20] I. Borutta, S. Sosnowski, K. Künlenz, M. Zehetleitner, and N. Bischof. Generating artificial smile variations based on a psychological system-theoretic approach. In *Proc. of the 18th IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN)*, Toyama, Japan, 2009.

[21] J. Boye and M. Wirén. "Multi-slot semantics for natural-language call routing systems. In *Proceedings of the Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies*, pages 68–75, 2007.

[22] Margaret M. Bradley and Peter J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49 – 59, 1994.

[23] C. Breazeal. Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, 59(1-2):119–155, 2003.

[24] C. Breazeal and C.D. Kidd. A robotic weight loss coach. In *Proceedings of the International Conference on Artificial Intelligence*, 2007.

[25] C. Breazeal and B. Scassellati. Infant-like social interactions between a robot and a human caregiver. *Adaptive Behavior*, 8(1):49–74, 2000.

[26] C. L. Breazeal. *Designing Sociable Robots*. MIT Press, 2001.

[27] J.S. Bruner and L. Postman. On the perception of incongruity: a paradigm. *Journal of Personality*, 18:206–223, 1949.

[28] M. Buss, D. Carton, B. Gonsior, K. Kühnlenz, C. Landsiedel, N. Mitsou, R. de Nijs, J. Zlotowski, S. Sosnowski, E. Strasser, M. Tscheligi, A. Weiss, and D. Wollherr. Towards proactive human-robot interaction in human environments. In *Proc. of the Int. Conf. on Cognitive Infocommunications (CogInfoCom)*, 2011.

[29] L. Canamero and J. Fredslund. I show you how i like you - can you read it in my face? *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 31(5):454–459, Sept. 2001.

[30] R. Carston. The explicit/implicit distinction in pragmatics and the limits of explicit communication. *International Review of Pragmatics*, 1(1):35–62, 2009.

[31] J. Cassell and T. Bickmore. Relational agents: A model and implementation of building user trust. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 396–403, 2001.

[32] C. Castelfranchi. Modelling social action for ai agents. *Artificial Intelligence*, 103:157–182, 1998.

[33] C. Castelfranchi. Grounding social action and phenomena in mental representations. In *Advances in Cognitive Science: Learning, Evolution, and Social Action. Proc. of IWCogSc-10- ILCLI*, pages 93–112, 2010.

[34] T. L. Chartrand and J. A. Bargh. The chameleon effect: the perception-behavior link and social interaction. *J Pers Soc Psychol*, 76(6):893–910, Jun 1999.

[35] J. Chu-Carroll and B. Carpenter. Vector-based natural language call routing. *Computational Linguistics*, 25(3):361–388, September 1999.

[36] H. H. Clark and E. F. Schaefer. Contributing to discourse. *Cognitive Science*, 13:259–294, 1989.

[37] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. Taylor. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18(1):32 – 80, 2001.

[38] H. Cramer, J. Goddijn, B. Wielinga, and V. Evers. Effects of (in)accurate empathy and situational valence on attitudes towards robots. In *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 141–142, New York, USA, 2010. ACM Press.

[39] O. Damm, K. Malchus, F. Hegel, P. Jaecks, P. Stenneken, B. Wrede, and M. Hielscher-Fastabend. A computational model of emotional alignment. In *Proc. of 5th Workshop on Emotion and Computing*, Berlin, 2011.

[40] M. Dapretto, M. S. Davies, J. H. Pfeifer, A. A. Scott, M. Sigman, S. Y. Bookheimer, and M. Iacoboni. Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience*, 9:28–30, 2005.

[41] K. Dautenhahn. Socially intelligent robots: dimensions of human-robot interaction. *Journal of Philosophical Interactions*, pages 679–704, 2007.

[42] F. Delaunay and T. Belpaeme. Refined human-robot interaction through retro-projected robotic heads. In *Proceedings of the IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*, pages 106–107, 2012.

[43] S. Eggins and D. Slade. *Analysing Casual Conversation*. Cassell, 1997.

[44] P. Ekman. Universals and cultural differences in facial expressions of emotion. In J. Cole, editor, *Proc. of the Symposium on Motivation*, volume 19, pages 207–283, University of Nebraska, 1971.

[45] P. Ekman and W.V. Friesen. *Investigator's Guide: Part two. Facial Action Coding System*. Consulting Psychologists Press, Palo Alto, CA, 1978.

[46] C. Elliott, J. Rickel, and J. Lester. *Artificial Intelligence Today: Recent Trends and Developments*, volume 1600 of *Lecture Notes in Computer Science*, chapter Lifelike Pedagogical Agents and Affective Computing: An Exploratory Synthesis, pages 195–211. Springer Berlin Heidelberg, 1999.

[47] J. Fink. Anthropomorphism and human likeness in the design of robots and human-robot interaction. In *Proceedings of the International Conference on Social Robotics (ICSR)*, 2012.

[48] J. Fink, O. Mubin, F. Kaplan, and P. Dillenbourg. Anthropomorphic language in online forums about roomba, aibo and the ipad. In *Proceedings of the IEEE International Workshop on Advanced Robotics and its Social Impacts (ARSO)*, pages 54–59, 2012.

[49] A. H. Fischer and G. A. van Kleef. Where have all the people gone? a plea for including social interaction in emotion research. *Emotion Review*, 2(3):208–211, 2010.

[50] B.J. Fogg. *Persuasive Technology: Using Computers to Change What We Think and Do*, chapter 5: Computers as Persuasive Social Actors, pages 89–120. The Morgan Kaufmann Series in Interactive Technologies. Morgan Kaufmann Publishers, San Francisco, 2003.

[51] D. Frey and M. Irle, editors. *Theorien der Sozialpsychologie, Band II: Gruppen-, Interaktions- und Lerntheorien*, volume 2. Verlag Hans Huber, 2002.

[52] M. Gabsdil. Clarification in spoken dialogue systems:. In *Proceedings of AAAI Spring Symposium Workshop on Natural Language Generation in Spoken and Written Dialogue*, 2003.

[53] V. Gallese. The 'shared manifold' hypothesis. *Journal of Consciousness Studies*, 8, 2001.

[54] D. Gentner. *International Encyclopedia of the Social and Behavioral Sciences*, chapter Psychology of Mental Models, pages 9683–9687. Elsevier, 2002.

[55] M. Goebl and G. Färber. A real-time-capable hard- and software architecture for joint image and knowledge processing in cognitive automobiles. *Intelligent Vehicles Symposium*, pages 737–740, 2007.

[56] B. Gonsior, M. Buß, S. Sosnowski, D. Wollherr, K. Kühnlenz, and M. Buss. Towards transferability of theories on prosocial behavior from social psychology to HRI. In *Proc. of the IEEE Int. Workshop on Advanced Robotics and its Social Impacts (ARSO)*, pages 101–103, Munich, 2012.

[57] B. Gonsior, C. Landsiedel, A. Glaser, D. Wollherr, and M. Buss. Dialog strategies for handling miscommunication in task-related HRI. In *Proc. of IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN)*, pages 369–375, Atlanta, GA, USA, 2011.

[58] B. Gonsior, C. Landsiedel, N. Mirnig, S. Sosnowski, E. Strasser, J. Zlotowski, M. Buss, K. Kühnlenz, M. Tscheligi, A. Weiss, and D. Wollherr. Impacts of multimodal feedback on efficiency of proactive information retrieval from task-related HRI. *Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII), Special Issue on Cognitive Infocommunications*, 16(2):313–326, 2012.

[59] B. Gonsior, S. Sosnowski, M. Buß, D. Wollherr, and K. Kühnlenz. An emotional adaption approach to increase helpfulness towards a robot. In *Proc. of the IEEE Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 2429–2436, Vilamoura, 2012.

[60] B. Gonsior, S. Sosnowski, C. Mayer, J. Blume, B. Radig, D. Wollherr, and K. Kühnlenz. Improving aspects of empathy and subjective performance for HRI through mirroring facial expressions. In *Proc. of IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN)*, pages 350–356, Atlanta, GA, USA, 2011.

[61] B. Gonsior, D. Wollherr, and M. Buss. Towards a dialog strategy for handling miscommunication in human-robot dialog. In *Proc. of IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN)*, pages 284–289, Viareggio, Italy, 2010.

[62] A. L. Gorin, G. Riccardi, and J.H. Wright. How may i help you? *Speech Communication*, 23:113–127, 1997.

[63] J. Gratch, N. Wang, J. Gerten, E. Fast, and R. Duffy. Creating rapport with virtual agents. In C.Pelachaud, J.-C. Martin, E. Andr, G. Chollet, K. Karpouzis, and D. Pel, editors, *Intelligent Virtual Agents*, volume 4722 of *Lecture Notes in Computer Science*, pages 125–138. Springer Berlin / Heidelberg, 2007.

[64] H. Gubler and N. Bischof. A systems' perspective on infant development. *Infant Development: Perspectives from German-speaking Countries*, pages 1–37, 1990.

[65] F. Hara, H. Kobayashi, F. Iida, and M. Tabata. Personality characterization of animate face robot through interactive communication with human. In *Proc. of Int. Workshop on Humanoid and Human Friendly Robotics (IARP)*, Tsukuba, Japan, 1998.

[66] M. Heerink, B. Krose, V. Evers, and B. Wielinga. Measuring acceptance of an assistive social robot: a suggested toolkit. In *Proc. of the 18th IEEE Int. Symposium on Robot and Human Interactive Communication, 2009. (RO-MAN)*, pages 528–533, 2009.

[67] F. Hegel, F. Eyssel, and B. Wrede. The social robot FLOBI: Key concepts of industrial design. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 107–112, 2010.

[68] D. R. Heise. *Agent culture: human-agent interaction in a multicultural world*, chapter 6: Enculturating agents with expressive role behavior, pages 127–142. Lawrence Erlbaum, 2004.

[69] U. Hess and S. Blairy. Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *Int. Journal of Psychophysiology*, 40(2):129 – 141, 2001.

[70] G. Hirst, S. McRoy, P. Heeman, P. Edmonds, and D. Horton. Repairing conversational misunderstandings and non-understandings. *Speech Communication*, 15(3-4):213–229, 1994.

[71] K. Isbister, C. Nass, H. Nakanishi, and T. Ishida. *Agent Culture: Human-Agent Interaction in a Multicultural World*, chapter Designing a Social Agent for Virtual Meeting Space, pages 245–266. Lawrence Erlbaum Associates, 2004.

[72] R. Kaliouby, R. Picard, and S. Baron-Cohen. *Annals of the New York Academy of Sciences*, volume 1093 of *Progress in Convergence*, chapter Affective Computing and Autism, pages 228–248. 2006.

[73] K. Kamei, K. Shinozawa, T. Ikeda, A. Utsumi, T. Miyashita, and N. Hagita. Recommendation from robots in a real-world retail shop. In *Workshop on Machine Learning for Multimodal Interaction, Multimodal Interfaces*, 2010.

[74] T. Kanda, M. Kamasima, M. Imai, T. Ono, D. Sakamoto, H. Ishiguro, and Y. Anzai. A humanoid robot that pretends to listen to route guidance from a human. *Autonomous Robots*, 22:87–100, 2007.

[75] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita. An affective guide robot in a shopping mall. In *Proceedings of HRI'09*, pages 173–180, 2009.

[76] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita. A communication robot in a shopping mall. *Robotics*, 26(5):897–913, 2010.

[77] M. Karg. *Pattern Recognition Algorithms for Gait Analysis with Application to Affective Computing*. PhD thesis, Institute of Automatic Control Engineering (LSR), Technische Universität München, 2012.

[78] J.F. Kelley. An empirical methodology for writing user-friendly natural language computer applications. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems (CHI)*, pages 193–196, 1983.

[79] S. Kiesler and J. Goetz. Mental models and cooperation with robotic assistants. In *Proc. of the ACM Conference on Human Factors in Computing Systems (CHI)*. ACM SIGCHI, 2006.

[80] K. H. Kim, S. W. Bang, and S. R. Kim. Emotion recognition system using short-term monitoring of physiological signals. *Medical and Biological Engineering and Computing*, 42(3):419–427, 2004.

[81] H. Kobayashi, F. Hara, and A. Tange. A basic study on dynamic control of facial expressions for face robot. In *Proceedings of the 3rd International Workshop on Robot and Human Communication (RO-MAN)*, pages 168–173, 1994.

[82] T. Koulouri and S. Lauria. Exploring miscommunication and collaborative behaviour in human-robot interaction. In *Proceedings of the SIGDIAL 2009 Conference*, pages 111–119, Morristown, NJ, USA, 2009. Association for Computational Linguistics.

[83] R. E. Kraut and R. E. Johnston. Social and emotional messages of smiling: An ethological approach. *Journal of Personality and Social Psychology*, 37(9):1539–1553, 1979.

[84] D. Krebs. Empathy and altruism. *Journal of Personality and Social Psychology*, 32:1134–1146, 1975.

[85] B. Kühnlenz, S. Sosnowski, M. Buß, D. Wollherr, K. Kühnlenz, and M. Buss. Increasing helpfulness towards a robot by emotional adaption to the user. *International Journal of Social Robotics (IJSR)*, pages 1–20. Springer, 2013.

[86] K. Kühnlenz, S. Sosnowski, and M. Buss. Impact of animal-like features on emotion expression of robot head eddie. *Advanced Robotics*, 24(8-9):1239–1255, 2010.

[87] V.A Kulyukin. On natural language dialogue with assistive robots. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, pages 164–171. ACM, 2006.

[88] T. Kuratate, Y. Matsusaka, B. Pierce, and G. Cheng. mask-bot: A life-size robot head using talking head animation for human-robot communication. In *Proceedings of the 11th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, pages 99–104, 2011.

[89] S. Lauria, G. Bugmann, T. Kyriacou, and E. Klein. Instruction Based Learning: How to instruct a personal robot to find HAL. In *European Workshop on Learning Robots*, 2001.

[90] H. Lee, J. Park, and M. Chung. A linear affectexpression space model and control points for mascot-type facial robots. *IEEE Transactions on Robotics*, 23(5):863–873, 2007.

[91] M.K. Lee, S. Kiesler, and J. Forlizzi. Receptionist or information kiosk: How do people talk with a robot? In *Proceedings of the ACM Conference on Computer supported Cooperative Work (CSCW)*, pages 31–40, New York, NY, USA, 2010. ACM.

[92] C. Lichtenthäler, T. Lorenz, M. Karg, and A. Kirsch. Increasing perceived value between human and robots - measuring legibility in human aware navigation. In *Proc. of the IEEE Int. Workshop on Advanced Robotics and its Social Impacts (ARSO)*, 2012.

[93] W. Lilli and D. Frey. *Theorien der Sozialpsychologie [Theories of Social Psychology]*, volume 1: Cognitive Theories. Hans Huber Verlag, Bern, 2 edition, 1993.

[94] C. Liu, K. Conn, N. Sarkar, and W. Stone. Online affect detection and robot behavior adaptation for intervention of children with autism. *IEEE Transactions on Robotics*, 24(4):883–896, 2008.

[95] R. Looijea, M. A. Neerincxa, and F. Cnossenc. Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. *International Journal of Human-Computer Studies*, 68(6):386–397, 2010.

[96] N. Mattar and I. Wachsmuth. Small talk is more than chit-chat. *Lecture Notes in Computer Science-KI 2012: Advances in Artificial Intelligence*, 7526:119–130, 2012.

[97] C. Mayer, S. Sosnowski, K. Kühnlenz, and B. Radig. Towards robotic facial mimicry: System development and evaluation. In *Proc. of the IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN)*, 2010.

[98] C. Mayer, M. Wimmer, M. Eggers, and B. Radig. Facial expression recognition with 3d deformable models. In *Proc. of the 2nd Int. Conf. on Advancements Computer-Human Interaction (ACHI)*. Springer, 2009.

[99] S.W. McQuiggan and J.C. Lester. Modeling and evaluating empathy in embodied companion agents. *International Journal of Human-Computer Studies*, 65(4):348–360, 2007.

[100] A. Mehrabian. *Silent Messages: Implicit Communication of Emotions and Attitudes*. Wadsworth, Belmont, CA, USA, 1981.

[101] A. Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament. *Current Psychology*, 14(4):261–292, 1996.

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

[103] N. Mirnig, B. Gonsior, S. Sosnowski, C. Landsiedel, D. Wollherr, A. Weiss, and M. Tscheligi. Feedback guidelines for multimodal human-robot interaction: How should a robot give feedback when asking for directions? In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 533–538, 2012.

[104] N. Mirnig, B. Gonsior, D. Wollherr, A. Weiss, and M. Tscheligi. Feedback in human-robot interaction: How to display a robots internal system status. In *Proceedings of the 3rd International Conference on Social Robotics (ICSR)*, 2011.

[105] S. Al Moubayed, J. Beskow, G. Skantze, and B. Granström. *Cognitive Behavioural Systems. Lecture Notes in Computer Science.*, chapter Furhat: A Back-Projected Human-Like Robot Head for Multiparty Human-Machine Interaction. Springer, 2012.

[106] R. Müller, T. Röfer, A. Lankenau, R. Musto, K. Stein, and A. Eisenkolb. Coarse qualitative descriptions in robot navigation. In *Spatial Cognition II*, pages 265–276, 2000.

[107] J. Mumm and B. Mutlu. Human-robot proxemics: physical and psychological distancing in human-robot interaction. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 331–338. IEEE, 2011.

[108] K. Nakagawa, M. Shiomi, K. Shinozawa, R. Matsumura, H. Ishiguro, and N. Hagita. Effect of robot's active touch on people's motivation. In *Proc. of IEEE Int. Conf. on Human-Robot Interaction (HRI)*, 2011.

[109] K. Nakagawa, M. Shiomi, K. Shinozawa, R. Matsumura, H. Ishiguro, and N. Hagita. Effect of robot's whispering behavior on people's motivation. *International Journal of Social Robotics*, 5(1):5–16, January 2013.

[110] R. Niewiadomski, M. Ochs, and C. Pelachaud. Expressions of empathy in ECAs. *Intelligent Virtual Agents*, pages 1–8, 2008.

[111] S. Nishio, H. Ishiguro, and N. Hagita. Geminoid: Teleoperated android of an existing person. *Humanoid Robots-New Developments*, 14, 2007.

[112] I. R. Nourbakhsh, J. Bobenage, S. Grange, R. Lutz, R. Meyer, and A. Soto. An affective mobile robot educator with a full-time job. *Artificial Intelligence*, 114(12):95 –124, 1999.

[113] M. Ochs, C. Pelachaud, and D. Sadek. An empathic virtual dialog agent to improve human-machine interaction. In Padgham, Parkes, Müller, and Parsons, editors, *Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 89–96, Estoril, Portugal, 2008.

[114] A. Paiva, J. Dias, D. Sobral, R. Aylett, S. Woods, L. Hall, and C. Zoll. Learning By Feeling: Evoking Empathy With Synthetic Characters. *Applied Artificial Intelligence*, 19(3-4):235–266, 2005.

[115] R.W. Picard. *Affective Computing*. MIT Press, Cambridge, 1997.

[116] M. J. Pickering and S. Garrod. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, pages 1–58, 2004.

[117] M. J. Pickering and S. Garrod. Alignment as the basis for successful communication. *Research on Language and Computation*, 4(2-3):203–228, 2006.

[118] H. Prendinger and M. Ishizuka. The emphatic companion: A character-based interface that adresses users' affective states. *Applied Artificial Intelligence*, 19(3-4):267–285, 2005.

[119] P. Rani, C. Liu, N. Sarkar, and E. Vanman. An empirical study of machine learning techniques for affect recognition in humanrobot interaction. *Pattern Analysis and Applications*, 9(1):58–69, 2006.

[120] B. Reeves and C. Nass. *The media equation: how people treat computers, television, and new media like real people and places*. Cambridge University Press, New York, NY, USA, 1998.

[121] W. Reitberger, A. Meschtscherjakov, T. Mirlacher, T. Scherndl, H. Huber, and M. Tscheligi. A persuasive interactive mannequin for shop windows. In *Proc. of the 4th Int. Conf. on Persuasive Technology (Persuasive)*. ACM, 2009.

[122] L. Riek, P. Paul, and P. Robinson. When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry. *Journal on Multimodal User Interfaces*, 3:99–108, 2010.

[123] L.D. Riek. Wizard of oz studies in hri: A systematic review and new reporting guidelines. *Journal of Human Robot Interaction*, 1(1):119–136, 2012.

[124] L.D. Riek and P. Robinson. Real-time empathy: Facial mimicry on a robot. In *in Workshop on Affective Interaction in Natural Environments (AFFINE) at the International ACM Conference on Multimodal Interfaces. ACM*, pages 1–5, 2008.

[125] R. E. Riggio, J. Tucker, and D. Coffaro. Social skills and empathy. *Personality and Individual Differences*, 10(1):93–99, 1989.

[126] S. Sabanovic, M.P. Michalowski, and R. Simmons. Robots in the wild: Observing human-robot social interaction outside the lab. In *IEEE International Workshop on Advanced Motion Control*, pages 596–601, 2006.

[127] P. Salvini, G. Ciaravella, W. Yu, G. Ferri, A. Manzi, B. Mazzolai, C. Laschi, S. R. Oh, and P. Dario. How safe are service robots in urban environments? bullying a robot. In *Proc. of IEEE International Symposium on Robot and Human Interactive Communication*, 2010.

[128] J. Satake and J. Miura. Multiple-person tracking for a mobile robot using stereo. In *Proceedings of MVA*, pages 273–277, 2009.

[129] M. Scheeff, J. Pinto, K. Rahardja, S. Snibbe, and R. Tow. Experiences with sparky, a social robot. *Socially Intelligent Agents*, 3:173–180, 2002.

[130] M. Schröder. The german text-to-speech synthesis system mary: A tool for research, development and teaching. pages 365–377, 2001.

[131] M. Schröder. Dimensional emotion representation as a basis for speech synthesis with non-extreme emotions. In *Proc. of Workshop on Affective Dialogue Systems*, pages 209–220, 2004.

[132] H. Shi and T. Tenbrink. Telling Rolland where to go: HRI dialogues on route navigation. In *WoSLaD Workshop on Spatial Language and Dialogue*, pages 23–25, 2005.

[133] H. Shibata, M. Kanoh, S. Kato, and H. Itoh. A system for converting robot emotion into facial expressions. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 3660–3665, 2006.

[134] M. Siegel, C. Breazeal, and M.I. Norton. Persuasive robotics: The influence of robot gender on human behavior. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2563–2568, 2009.

[135] G. Skantze. Exploring human error recovery strategies: Implications for spoken dialogue systems. *Speech Communication*, 45(3):325–341, 2005. Special Issue on Error Handling in Spoken Dialogue Systems.

[136] G. Skantze and S. Al Moubayed. Iristk: a statechart-based toolkit for multi-party face-to-face interaction. In *Proceedings of ICMI*, Santa Monica, CA., 2012.

[137] Gabriel Skantze. *Error Handling in Spoken Dialogue Systems - Managing Uncertainty, Grounding and Miscommunication*. Doctoral thesis in speech communication, KTH Royal Institute of Technology, Stockholm, Sweden, 2007.

[138] J.O. Smith. *Spectral Audio Signal Processing*. W3K Publishing, 2011.

[139] S. Sosnowski, K. Kuehnlenz, and M. Buss. Eddie - an emotion display with dynamic intuitive expressions. In *Proc. IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN)*, 2006.

[140] S. Sosnowski, C. Mayer, B. Radig, and K. Kühnlenz. Mirror my emotions! combining facial expression analysis and synthesis on a robot. In *Proc. of the 36th Annual Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB)*, 2010.

[141] R. A. Spitz. *No and yes: on the genesis of human communication*. International Universities Press, 1957.

[142] R. N. Spreng, M.C. McKinnon, R.A. Mar, and B. Levine. The toronto empathy questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. *Journal of Personality Assessment*, 91(1):62–71, 2009.

[143] T. Stocky, J.Cassell, Y. I. Nakano, and G.Reinstein. Towards a model of face-to-face grounding. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 553–561, 2003.

[144] L. A. Suchman. *Human-machine reconfigurations: Plans and situated actions*. Cambridge University Press, Cambridge, UK, 2007.

[145] L.A. Suchman. *Plans and situated actions: The problem of human-machine communications*. Cambridge University Press, Cambridge, UK, 1987.

[146] L.A. Suchman. Reconfiguring human-robot relations. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 652–654, 2006.

[147] B.G. Tabachnick and L.S. Fidell. *Experimental Design Using ANOVA*. Duxbury Applied Series. Brooks/Cole, 2007.

[148] A. Takanishi, K. Sato, K. Segawa, H. Takanobu, and H. Miwa. An anthropomorphic head-eye robot expressing emotions based on equations of emotion. In *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, pages 2243–2249, 2000.

[149] A. Tapus and M. J Mataric'. Emulating Empathy in Socially Assistive Robotics Empathy in Socially Assistive Robotics. In *In AAAI Spring Symposium on Multidisciplinary Collaboration for Socially Assistive Robotics*, Palo Alto, Stanford, U.S.A., 2007.

[150] A. P. Thomas, P. Bull, and D. Roger. Conversational exchange analysis. *Journal of Language and Social Psychology*, 1(2):141–156, 1982.

[151] A. Thomaz, M. Berlin, and C. Breazeal. An embodied computational model of social referencing. In *Proceedings of the IEEE International Workshop on Robot and Human In-teractive Communication (RO-MAN)*, pages 591–598, 2005.

[152] D. R. Traum. *A Computational Theory of Grounding in Natural Language Conversation*. PhD thesis, University of Rochester, Computer Science, 1994.

[153] A.J.N. van Breemen, K. Crucq, B.J.A Kröse, M. Nuttin, J.M. Porta, and E. Demeester. A user-interface robot for ambient intelligent environments. In *Proceedings of the 1st International Workshop on Advances in Service Robotics (ASER)*, pages 132–139. Fraunhofer IRB Verlag, 2003.

[154] A.J.N. van Breemen, X. Yan, and B. Meerbeek. iCat: An animated user-interface robot with personality. In *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems (ACM)*, pages 143–144, 2005.

[155] J.M. van der Zwaan, V. Dignum, and C.M. Jonker. A BDI Dialogue Agent for Social Support : Specification of Verbal Support Types (Extended Abstract) Categories and Subject Descriptors. In Conitzer, Winikoff, Padgham, and van der Hoek, editors, *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, Valencia, Spain, 2012.

[156] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis. User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27:425–478, 2003.

[157] E. Vlacho and H. Schärfe. Android emotions revealed. In *Proceedings of the 4th international conference on Social Robotics (ICSR)*, pages 56–65. Springer, 2012.

[158] R. F. Voss and J. Clarke. 1/f noise in music: Music from 1/f noise. *Journal of the Acoustical Society of America (JASA)*, 63(1):258–263, 1978.

[159] L. Vygotskij. *Thought and Language*. The M.I.T. Press,, 1962.

[160] K. Wada and T. Shibata. Living with seal robots - its sociopsychological and physiological influences on the elderly at a care house. *IEEE Transactions on Robotics*, 23(5):972–980, 2007.

[161] F. Wallhoff, T. Rehrl, C. Mayer, and B. Radig. Realtime face and gesture analysis for human-robot interaction. In *Proc. of the SPIE, Society of Photo-Optical Instrumentation Engineers Conf.*, 2010.

[162] J. Wang, D. Busquets, R. Gockley, R. Simmons, D. Busquets, C. Di Salvo, K. Caffrey, S. Rosenthal, J. Mink, S. Thomas, W. Adams, T. Lauducci, M. Bugajska, D. Perzanowski, and A. Schultz. Grace and george: Social robots at AAAI. In *Proceedings of AAAI'04. Mobile Robot Competition Workshop (Technical Report WS-04-11)*, 2004.

[163] P. Watzlawick, J. H. Beavin, and D. D. Jackson. *Pragmatics of Human Communication.* New York, Norton, 1967.

[164] A. Weiss, R. Bernhaupt, M. Tscheligi, D. Wollherr, K. Kühnlenz, and M. Buss. A methodological variation for acceptance evaluation of human-robot interaction in public places. In *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 713–718, 2008.

[165] A. Weiss, J. Igelsböck, M. Tscheligi, A. Bauer, K. Kühnlenz, D. Wollherr, and M. Buss. Robots asking for directions: The willingness of passers-by to support robots. In *Int. Conf. on Human-Robot Interaction (HRI)*, pages 23–30, 2010.

[166] S. Werner, B. Krieg-Brückner, and T. Herrmann. Modelling Navigational Knowledge by Route Graphs. *Spatial cognition II*, pages 295–316, 2000.

[167] W. Wirth, T. Hartmann, S. Bcking, P. Vorderer, C. Klimmt, and H. et al. Schramm. A process model of the formation of spatial presence experiences. *Journal of Media Psychology*, 9:493–525, 2007.

[168] A. Wojdel and L. J. M. Rothkrantz. Facs based generating of facial expressions. In *Proc. of 7th Annual Conf. of the Advanced School for Computing and Imaging (ASCI)*, 2001.

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

[170] Y. Yamaji, T. Miyake, Y. Yoshiike, P.R. De Silva, and M. Okada. Stb: human-dependent sociable trash box. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 197–198, 2010.

[171] J. E. Young, R. Hawkins, E. Sharlin, and T. Igarashi. Toward acceptable domestic robots: Apllying insights from social psychology. *Int. Journal of Social Robotics (IJSR)*, 1(1):95–108, 2009.

[172] M. Zecca, S. Roccella, M. C. Carrozza, H. Miwa, K. Itoh, G. Cappiello, J.-J. Cabibihan, M. Matsumoto, H. Takanobu, P. Dario, and A. Takanishi. On the development of the emotion expression humanoid robot WE-4RII with RCH-1. In S. Roccella, editor, *Proc. of 4th IEEE/RAS Int. Conf. on Humanoid Robots (Humanoids)*, pages 235–252, 2004.