Paper: jc*--**-****

Impacts of Multimodal Feedback on Efficiency of Proactive Information Retrieval from Task-Related HRI

Barbara Gonsior, Christian Landsiedel, Nicole Mirnig, Stefan Sosnowski, Ewald Strasser, Jakub Złotowski
Martin Buss, Kolja Kühnlenz, Manfred Tscheligi, Astrid Weiss, Dirk Wollherr

Institute of Automatic Control Engineering (LSR), Technische Universität München, D-80290 Munich, Germany
ICT&S Center, University of Salzburg, Sigmund-Haffner-Gasse 18, 5020 Salzburg, Austria
E-mail: bg@tum.de
[Received 00/00/00; accepted 00/00/00]

Abstract. This work is a first step towards an integration of multimodality with the aim to make efficient use of both human-like, and non-human-like feedback modalities in order to optimize proactive information retrieval from task-related Human-Robot Interaction (HRI) in human environments. The presented approach combines the human-like modalities speech and emotional facial mimicry with non-human-like modalities. The proposed non-human-like modalities are a screen displaying retrieved knowledge of the robot to the human and a pointer mounted above the robot head for pointing directions and referring to objects in shared visual space as an equivalent for arm and hand gestures. Initially, pre-interaction feedback is explored in an experiment investigating different approach behaviors in order to find socially acceptable trajectories to increase the success of interactions and thus efficiency of information retrieval. Secondly, pre-evaluated human-like modalities are introduced. First results of a multimodal feedback study are presented in the context of the IURO project[1], where a robot asks for its way to a predefined goal location.

Keywords: robotics, human-robot interaction, emotions, facial expressions, usability

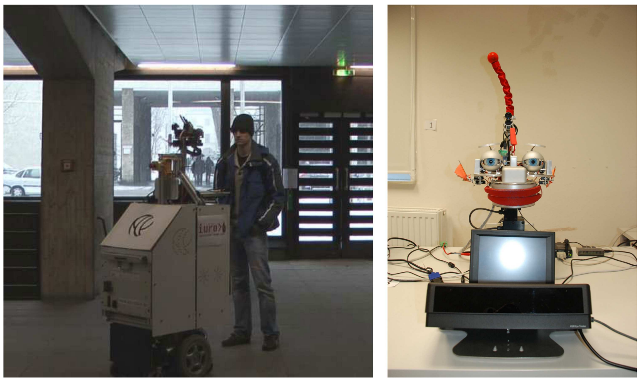

Fig. 1. Snapshot from the video presented to the participants (left); experimental setup with pointer, EDDIE head, touchscreen and eye tracker (right).

1. Introduction

Nowadays, robots are moving from highly controlled and strictly defined laboratory settings to natural human environments. Museum guides [35, 49] and shopping assistants [20] for indoor environments and surveillance robots [39] or garbage collectors [12] for outdoor scenarios are some of the applications that have evolved towards this end. As a consequence, accessible schemes for HRI gain more importance. Accordingly, natural, robust and interactive user interfaces are needed to enable users to successfully and efficiently interact with robots.

The use and combination of different modalities in HRI are necessary for creating a more complete interaction experience [11]. An example for combining speech, pointing gestures and head orientation in a multimodal human-robot dialog system is described in [47]. Providing feedback by means of multimodal interaction channels is an important factor for the success of conversations, not only for human-human interaction [32], but also for HRI, since the internal system status of a robot can be displayed to the human interaction partner, and thus helps to interpret and to understand the utterances and actions of a robot [31]. Since robots are not limited to the set of human modalities, research on different multimodal variations of feedback [38] provide a valuable contribution.

In this article, we adress proactive HRI in terms of necessary specifications for various cognitive feedback channels to the user, consisting of pre-interaction approaching behaviors, dialog strategies, emotional feedback as well as graphical and pointing interfaces. More specifically, we discuss issues related to the operation of robots in urban, public spaces as occurring in the IURO project. There, an autonomous robot is given the task of finding a pre-defined goal location in a city without map knowledge, purely relying on direction information gathered from the interaction with passers-by. Thus, for the experiments in this paper, the route description domain was

1. Interactive Urban Robot, `http://www.iuro-project.eu`

chosen as it provides a challenging but valid and rather well-explored stucture for interaction. Nevertheless, the approach is applicable to any proactive task-related information retrieval by a robot from humans.

In general, research on proactive behavior in robotics has focused on human intention recognition in order to make autonomous decisions on what task to execute [33, 42, 43]. Proactivity has also been studied in the context of domestic assistive robots in [8], where a clear distinction is drawn between on-demand assistance and proactive assistance. These studies have reported positive user evaluation of proactive robot behaviors. However, in these scenarios the human was the main beneficiary of the interaction process. In this article, we present techniques that are oriented towards scenarios where the beneficiary of the interaction is the robot. Since there is no a priori interest for communicating on the part of the human, such situations put additional strain on the interaction capabilities of the robot, as a successful communication channel must be created while nurturing the human interest level.

Robots proactively asking humans for directions in order to extract information about their environments are still operating in very restricted and well-structured indoor environments. In [34], coarse qualitative route descriptions are given to a wheelchair robot that navigates in an office floor. The office robot Jijo-2 [1] can learn locations of offices and staff by moving around and asking humans for information. A robot asking for the way at a conference site is presented in [30], and a miniature robot finds its way in a model town by asking for directions in [28]. 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 are required, as well as proper environment modeling for situatedness of the dialog.

As the IURO project focuses on outdoor scenarios there is no control over the environmental conditions, which may have great influence on speech recognition performance. Hence, speech recognition errors may occur frequently, and miscommunication has to be handled.

There are existing dialog models designed for handling miscommuncation, e.g. like incorporated in the spoken dialog system HIGGINS [45], and a dialog model using requesting subdialogs to solve communication problems [48]. However, our approach focuses on populated outdoor settings, where a robot is faced with a dynamically changing environment which results in varying speech recognition performance. This requires more flexibility for dialog strategies and handling requests which can be adapted to situational impacts.

Another aspect of importance for the design of a robot proactively starting interaction with humans is the way in which the robot attracts the attention of humans and how it approaches interaction candidates in a socially acceptable manner. An evaluation of a trajectory-based method for approaching humans is presented.

The goal of this article is to present an overview of aspects of HRI which are identified in the IURO project as being essential for robots autonomously carrying out tasks in populated environments with incomplete task knowledge. First evaluation results of a multimodal feedback study with specific focus on efficiency are presented with the aim of exploring different feedback modalities in this context. Based on previously explored single modalities, project related progress is outlined towards a multimodal approach to control a dialog in terms of proactive retrieval of missing task-knowledge [7].

The paper is organized as follows: In Section 2, pre-interaction feedback for human approach is introduced and evaluation results for approach behavior as feedback modality are presented. Section 3 describes different dialog strategies for proactive information retrieval from humans by natural language dialog with regard to efficiency and user experience. In Section 4, the role of facial expression feedback is described with a special focus on smiling as socially motivated signal in HRI. Section 5 provides an overview on the first results of a multimodal feedback study and a technical system where all previously introduced modalities are combined. Conclusions are given in Section 6.

2. Exploring pre-interaction feedback for human approach

When two strangers create an interaction space and become co-participants in a situation, a pre-beginning sequence [41] of interaction exists. A robot initiating an interaction in an acceptable manner in this phase is expected to achieve more successful interaction in less time resulting in higher efficiency. In this section, the acceptance of proactive approach behavior of a robot as a means of increasing efficiency of information retrieval is addressed.

In HRI, robot approach behavior is a recent active research field. In fetch-and-carry tasks, seated people prefer an approach from the frontal left or right direction rather than frontally or from behind. Mediating factors include the location in a room or the interaction partner being seated or standing, e.g. for persons standing in the middle of a room with sufficient space to move aside, frontal approach is acceptable [10, 50]. Previous experience with robots [26] or gender [10] also influence how people would like to be approached. These findings emphasize the high context dependence of approach behavior preferences and a need for validation in a public space. In [40], a model for approaching people in a shopping mall is developed, successfully increasing the robot's performance in initiating interactions. However, the focus is on the robot's performance while the comfort of people who were approached frontally is neglected.

Works in social psychology show that people have preferences regarding comfortable distances to interaction partners [21, 22]. These social spaces are influenced by intimacy, social standing and status of interaction partners. Works in HRI show that people accept closer distances to a robot (within personal (0.45-1.2 m) or even intimate space (0.15-0.45 m)) than with human interaction partners (social space (1.2-3.6 m)) [51, 52].

In this work, it is investigated how a robot in need of help should approach walking or standing pedestrians in a street. It is expected that a standing person approached in a public space will have the same preference as in [50] and a walking person will prefer to be approached either from frontal left or right directions rather than frontally to avoid collisions. Furthermore, preferred moving speeds and interaction distances are investigated.

A video-based [54, 55] study and an online survey are conducted, which are described in the following.

2.1. Video based study - Setup

Six short videos (approx. 15 seconds) are recorded with IURO approaching an actor in a university corridor and asking him for directions to the assembly hall. The robot approached either from the (1) frontal right, (2) frontal left or (3) frontally. The actor was either (a) standing or (b) walking towards the robot. Robot approaching speed is 0.6 m/s, slowing down to 0.4 m/s 2 meters in front of the actor and finally stopping within human personal space (0.45-1.2 m).

The video-based study was conducted in a laboratory with 30 participants recruited via the online blackboard of the Students' Union (ÖH) of Salzburg. Three videos are presented to each participant in the static (actor standing) or dynamic (actor walking) condition, see Fig. 1. After each video, the participants answer three questions on a 9-point Likert scale regarding preferences for the robot's approach behavior, whether participants

- liked the approach trajectory of IURO (1 - strongly disagree, 9 - strongly agree)
- thought the speed of IURO was appropriate (1 - too slow, 5 - about right, 9 - too fast)
- thought the stopping distance of IURO was appropriate (1 - too close, 5 - about right, 9 - too far)

In order to evaluate whether approaching a walking or standing person would result in differences regarding the unconscious attitude towards the robot and how the various feedback modalities (speed, distance and approach direction) are subliminally perceived by the participants, the Affect Misattribution Procedure (AMP) [37] is applied for each participant after three videos.

Remark: AMP measures implicit attitudes towards an object. Participants are instructed to rate whether they perceive abstract patterns, such as Chinese characters, as beautiful or ugly. Just before a Chinese character appears, another object, in this study the actor or robot, is presented to the participants. If they have a positive attitude towards the object, a Chinese character following it should also be perceived more favorably.

2.2. Video based study - Results

Due to the bimodal distribution of the data, a non-parametric Friedman's analysis of variance by ranks is conducted. In the static condition, a significant difference between mean rankings for the approach trajectories is noted (χ^2 (2) = 6.35, p = .04). The participants prefer the frontal right (mean rank χ = 2.3) and frontal left (χ = 2) trajectories over the frontal (χ = 1.7). However, post-hoc tests only show a non-significant trend for preferring the frontal right over frontal trajectory (z = 2.32, p = .06). In the dynamic condition no statistically significant differences for the approach trajectories are observed.

Furthermore, no significant differences are found between dynamic and static conditions regarding the robot's approach speed. The participants rated the proposed speed profile as appropriate (M = 5.09, SD = .55) and people felt comfortable with IURO stopping in human personal space (M = 4.5, SD = .79). However, a significant difference in the dynamic condition regarding the comfortable interaction distance (χ^2 (2) = 7.79, p = .02) is noted. Pairwise comparisons do not show significant differences between these conditions. However, a tendency for feeling that the robot stops too close in the frontal (mean rank χ = 1.64) than in the frontal left (χ = 2.19) or frontal right (χ = 2.17) trajectories is noted.

Finally, the results of the AMP show that the robot approaching a standing (M = 30.75, SD = 3.41) compared to a walking person (M = 28.32, SD = 4.07) does not affect the implicit attitudes of the participant towards it. However, analyzing the data from the participants who neither had prior experience with robots nor extreme conscious attitude towards them (based on a demographic questionnaire) shows a significant difference (t(20) = 2.11, p = .048). Therefore, people whose opinions can be influenced by a short video of HRI unconsciously perceive the robot more negatively when it approaches a walking (M = 28.19, SD = 4.08) compared to a standing person (M = 31.63, SD = 3.56).

2.3. Online survey - Setup

Since some non-significant trends were noted in the first study, a follow-up online survey with the same videos, but larger sample size was conducted [57]. A number of 182 participants were recruited using university and scientific mailing lists. The participants were randomly assigned to either the dynamic or static condition and were presented three videos each. After each video the participants were asked to answer the same three questions regarding the robot's approach behavior on a 5-point Likert scale (reduced from 9-point based on participants' feedback in the first study).

2.4. Online survey - Results

A repeated-measures ANOVA with Least Significant Difference (LSD) post-hocs for static and dynamic conditions is conducted. In the dynamic condition, Mauchly's test indicates that the assumption of sphericity is violated and a Greenhouse-Geisser correction is applied. A statistically significant effect in preferences for the robot's approach trajectories is noted F (1.71, 148.68) = 4.35, p = .02. Post-hoc comparisons show that participants prefer the robot approaching a walking person from the frontal left (M = 3.46, SD = .92) over frontally (M = 3.18, SD = .98) p = .049 and frontal right (M = 3.51, SD = .8) over frontally p = .01.

Moreover, the results confirm the non-significant trend observed in the first study for the robot's stopping distance F (1.42, 123.57) = 47.87, p = .00. Post-hoc comparisons show that the participants rated the interaction distance as too close, when the robot approaches frontally (M = 2.31, SD = .88) compared to the frontal left (M = 3.0, SD = .4) p = .00 and frontally compared to the frontal right (M = 3.07, SD = .54) p = .00. It is possible that a person walking towards a robot on a collision trajectory fears that an unknown robot will not stop and therefore finds the personal space too close to react. In the same situation when a robot approaches from either side, stopping less than 1.2 m is acceptable for people as they have more space to react and can avoid an accident by continuing to walk. In the static condition, people express no preferences regarding the robot's approach trajectories.

2.5. Approach behavior as feedback modality

The results of the two studies emphasize the importance of bodily feedback in order to communicate a robot's need for human help in the pre-opening phase of the interaction. A robot can increase its chances of receiving the required information and at the same time become more efficient by pro-actively approaching a human in a socially acceptable way. We found that approach trajectories can be an important feedback modality for robots seeking help. Approaching a walking person can lead to more negative implicit attitude towards the robot. Furthermore, in this situation it should avoid approaching frontally. When approaching a standing person in public space, all the approach trajectories are equally appropriate. In addition, distance can also be used by a robot to communicate its status and the robot should stop outside of the personal space (< 1.2 m) when approaching a walking person frontally in order not to scare them. Finally, we found that the used approach speed profile is appropriate for a robot requesting attention.

3. Natural Language Dialog

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 route retrieval task in the urban setting, the use of natural language dialog is also justified according to the modality selection formulated by Kulyukin [27], 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 information retrieved from humans for fulfilling its task. However, the use of natural language for human-robot interaction also entails a number of difficulties, e.g. vagueness and ambiguity of spoken language itself, and the technical challenges posed by automatic speech recognition.

Since speech recognition in outdoor environments is highly subject to errors, resulting misrecognition may mislead the dialog into wrong directions. Additionally, miscommunication may be caused by lexical or conceptual difficulties. Hence, it is important to develop dialog strategies for a robot in order to control the dialog structure and thus trigger suitable information input for the robot, and to deliver a natural and intuitive experience to the user. In order to achieve this, another important aspect is to identify and differentiate between several potential types of miscommunication [17]. Successful information retrieval and naturalness can be balanced by applying targeted handling requests in a flexible way independently from the dialog structure [16].

In the following, four different dialog strategies are presented. They allow adaption to inaccurate and unstable automatic speech recognition resulting from dynamically changing environmental impacts.

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 wheelchair-robot without any verbal feedback by the robot while driving around in a building. In a second step, a conceptual route graph is deduced that serves as a basis for later route inquiry dialogs [44]. When designing such dialogs, current approaches either apply open requests, e.g. "how may I help you?", and then classify the recognized speech input by means of machine learning [9, 18]. Another way is to use a dialog strategy, where the system asks closed questions, e.g. "should I head in this direction?" and thus breaks the task down into subtasks in order to increase solely task-specific information content with every question [6].

However, 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 create different dialog strategies in order to enable the robot to adapt to varying environmental conditions for speech recognition. These strategies are based on a common task-dependent information representation. For the route-direction task, this consists of directions and landmarks, i.e. salient places or objects the robot has to pass or at which the robot has to change directions.

A common structure of four consecutive phases is identified by analyzing asking-for-directions dialogs [56], see Fig. 2.

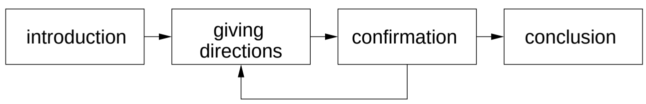

Fig. 2. Common structure of asking-for-directions dialog [56]

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. From Open Dialog to Closed Dialog, restrictedness is gradually increased to gain more structured and thus more predictable dialog behavior. Please note that these strategies are exemplarily confined to the context of asking passers-by for directions but applicable to any task-oriented inquiry dialog.

Open Dialog This strategy exactly applies the basic structure of human inquiry dialogs and thus should be the most intuitive one for humans: The robot opens the dialog in *introduction* phase by introducing itself and asking for its way to a certain goal location. After the passer-by has given route instructions during *giving directions* phase, the robot initializes *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 speak more clearly and to give the directions again in *giving directions* phase which can recur. 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 *introduction* phase. Yet, in *giving directions* phase the robot asks directly for separate route segments by proactively opening this phase with ”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 during *confirmation* phase in order to be confirmed or corrected by the human 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 the other strategies regarding its flow: A user cannot give any free information input, but is asked to confirm or revise mainly 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.

Summing up, all dialog strategies incorporate the above-mentioned basic structure, but differ in *giving directions* phase by allowing free speech input in Open Dialog, but 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 seperately in Divided, Requesting Divided, and Closed Dialog without the need of repeating the whole route as in Open Dialog.

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 [13] to different categories of miscommunication [24]. The resulting types of miscommunication and corresponding handling requests have been assigned to each state of the dialog, and then were combined with different dialog strategies in previous work [16].

3.2. Handling of Miscommunication

This Section provides an overview of requests for handling miscommunication which were implemented and combined with the above mentioned dialog strategies in a Wizard-of-Oz (WOz) experiment [16] and will be summarized with regard to efficiency in Sec. 3.3. A 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 subjects interact with an apparently autonomous computer system but which is actually being operated completely or in parts by an unseen human being (the ”wizard”) [25].

In the proposed approach, route descriptions given by the subjects are stored and processed internally based on route graphs [53], 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 experiment was 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 were 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 [13], 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 as explained more detailed in [17].

Reprise sluices mark the interpretation gap by emphasizing, e.g. ”wh”-alliterated words: *”Sorry,...*

textitwhere?/ *when?*/ *how far?*”

Wh-substituted reprises repeat the well-understood part and substitute interpretation 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 needed vocabulary for 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?*/ *break or fake?*”

”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 in the dialog phase of *confirmation*, after the robot has reconstructed the route description, correction requests are employed to revise an interpretation in order to determine the underlying intention of the speaker [13]: *”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.*”

3.3. Efficiency of Dialog Strategies and Handling Requests as Verbal Feedback Modalities

As mentioned above, the presented dialog strategies are evaluated in two different experiments [16].

Fully Automatic Indoor Experiment In the first experiment, the different dialog strategies are each modeled as state sequences according to their specifications given in Sec. 3.1, and were used as templates for a fully automatic (FA) indoor experiment, where a route displayed on a map was presented to the subjects, who were asked to describe the route to a spoken dialog system using DialogOS[2]. Four different maps were used in combination with each dialog strategy in randomized order.

Wizard-of-Oz Outdoor Experiment In the follow-up experiment, the different dialog strategies were combined with a variety of handling requests for miscommunication, which were employed in a WOz-experiment. In this experiment, for each of the dialog strategies, the subjects were engaged in dialogs with a robot platform with human-like features. The participants were asked to describe the best way to a wellknown, nearby location without the chance to consult a map. The different dialog strategies were modeled as finite state machines, where each state is either an input or an output state. In output nodes, speech output is performed using a text-to-speech system. The utterances are generated from templates that are filled with information from the system knowledge if necessary, e.g. in dialog states where information provided by a user is repeated to be verified.

In input states, information provided by the user is entered into the internal state of the system by the wizard

2. CLT Sprachtechnologie GmbH, www.clt-st.de

via a graphical interface. Thus, the wizard replaces a speech recognition component. Her actions are regulated by the way task knowledge is modeled similarily to the state-dependent grammars of a speech recognition system. Hence, the progression of the dialog as specified by the dialog strategy is enforced. For example, if a question related to a specific landmark in a route segment has been asked, the wizard is only able to refine the corresponding piece of information. After all input states in which relevant task-related information had been provided by the user, the wizard has the possibility to choose from different handling requests to simulate miscommunication.

The transitions between dialog nodes determining the course of the dialog are specified with conditions on existing knowledge and the progression of the dialog (e.g. requirements on information given as response to a question, such as posing a more refined question when not all the information requested has been provided in the answer of the user).

In order to provide a realistic setting for the dialog, the experiment took place in an outdoor environment at Technische Universität München, and the subjects were faced with the IURO robot as interaction partner.

In both experiments, participants filled in a questionnaire on their impression of the dialog using five-item Likert scales. In order to evaluate efficiency, the duration of interaction was calculated for each dialog.

Duration of Interaction and User Experience
The results of both experiments are presented in detail in [16], but can be summarized with regard to efficiency as follows: In comparison to the fully automatic experiment, where Closed Dialog was rated best and Open Dialog worst, the WOz-experiment revealed raised total scores for user experience with the best rating for Open Dialog strategy in combination with least use of miscommunication handling requests, and worst user ratings for Closed Dialog strategy in combination with the highest use of handling requests. As displayed in Fig. 3, the interaction length increased in the WOz-experiments compared to those in the fully automated experiment, except for Open Dialog, where least handling requests were applied. Additionally, user ratings of the dialog strategies in both experiments indicate correlations between user satisfaction, dialog duration, and the use of miscommunication handling requests. The vice-versa relation of user ratings for Open- and Closed Dialog strategy may be due to the fact that the wizard showed perfect speech recognition performance despite inducing miscommunication by the handling requests.

In summary, 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 is reversed decreasing user ratings again. Another factor of influence is the duration of the interaction: Within the WOz-experiment the duration of the Open Dialog condition is significantly shorter than all other dialog strategies and accordingly rated as most convenient

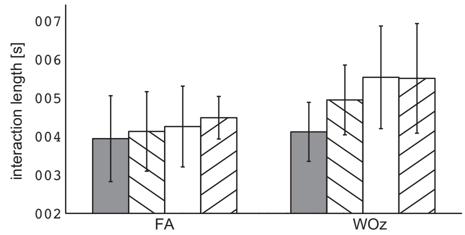

Fig. 3. Means and error bars (± standard deviation) for duration of dialogs of Fully Automatic (FA) and Wizard-of-Oz (WOz) scenario for all dialog strategies: (left to right) *Open*, *Divided*, *Requesting Divided* and *Closed Dialog*.

duration in the questionnaire. Due to the resulting significant increase of the total scores for Open Dialog compared between fully automated- and WOz-experiment 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.

These results are strongly motivating multimodality, since the number of used handling requests can be limited by multimodal feedback which may provide additional channels for preventing and handling miscommunication by communicating the internal state of the robot to the user. The following Section introduces facial expressions as one of these additional modalities.

4. Facial Expressions

Facial expressions are a way of communicating internal states of a robot and exchange social cues with interaction partners. In a setting like IURO, where untrained users interact with an unfamiliar robot, these can be used as intuitive reactions to guide the interaction, confirmation or disapproval of actions, or make the dialog more pleasant.

In the current system, the underlying model for the generation of facial expressions is based on the Facial Action Coding System (FACS) [36]. This allows an objective transfer of facial muscle movements, of which the corresponding expressions are composed, to the robotic face. EDDIE [46], a robot head developed for IURO combining anthropomorphic and zoomorphic features for displaying emotional expressions, is used as interaction partner (see Fig. 1 (right)).

EDDIE uses a subset of FACS including 13 action units (AUs) relevant for emotional expressions. The output of the previously described social motivation model directly controls the AUs to generate facial expressions. Visualization of speech is accomplished by parsing articulation information and generating a set of visemes to accompany speech output.

FACS also allows decoding of the facial expression of an interaction partner by extracting and processing features from camera images. The estimation of the activation of AUs in the human face is computed by a model-based approach measuring the deformation of a 3D facial model (Candide-III) fitted to the user's face [29]. With this combination of facial expression analysis and synthesis, IURO is able to react on visual emotional stimuli and provide a closed-loop interaction. An example of such an interaction is given in the following showing that IURO can utilize smiling as a form of socially relevant exchange of expressions.

4.1. Artificial Smiling

Smiling is not only a way to show affect, but also a socially motivated signal in interactions. It can give the interaction partner an impression of higher task performance, e.g. in scenarios of a server and a customer, and benefits from the perceived authenticity of the smile [19].

To artificially generate smiles in interactions with IURO, a system-theoretic approach based on a reduced version of the Zurich Model of Social Motivation [4] is used [5]. The model accounts for the motivational states of security, arousal and autonomy, which are homeostatic. A main assumption in this model is that smile reactions are a result of a decline in autonomy, meaning that smiles are a reaction on external disturbances of the homeostasis, like social distance changes, environmental changes, conflicts, etc. Changes in the respective subsystems lead to characteristic facial expressions, which in superposition result in the overall facial expression and distinctive smile variation. Parameters of external disturbances are derived from the camera images, extracting the distance to the human partner and his/her facial expression. In particular, the distance is a measure of the magnitude of the effects, meaning that signals from a person have lower impact, the further away that person is. Smiles are interpreted as an assurance of security, boosting the robot's security and thus provoking a smile reaction. Angry looks challenge the autonomy and afraid or surprised expressions increase the robot's arousal.

This setup enables the robot to react emotionally to users and their shown emotional states. By expressing its internal state via facial expressions, the user is able to comprehend the robot's state and react to it as well.

4.2. Artificial Smiling as Feedback Modality

The effect of the artificially generated smiles and emotional reactions are shown in a recent experiment [15] providing a higher rating in subjective performance and more empathy towards the robot in comparison to an emotionally neutral behaving robot. In this experiment, the robot is set up to a communicative task playing Akinator (www.akinator.com), a game of guessing a thought-of person by asking questions, with the robot as the guesser. The different conditions of neutral behavior and socially motivated reactions are evaluated towards user acceptance [23] and key concepts of HRI: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety [2]. Results are shown in Figs. 4 and 5.Additionally, two new measures for empathy and subjective performance are introduced and evaluated towards their impact on user acceptance. The hypothesis is that the condition with the robot showing emotional reactions outperforms the neutral condition in terms of user acceptance, empathy and subjective performance. In general, results support the hypothesis by showing a trend towards a better rating of the social motivation model being rated better in most instances than the neutral condition.

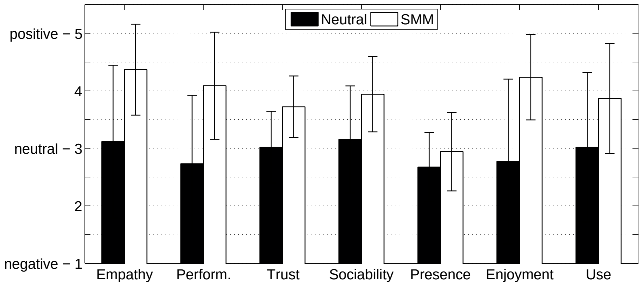

Fig. 4. Evaluation of user acceptance of the neutral behavior condition versus the socially motivated model

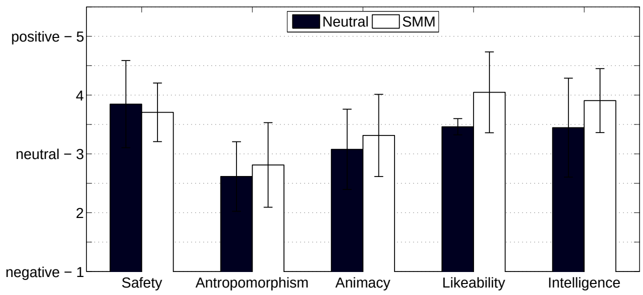

Fig. 5. Evaluation of key concepts of HRI of the neutral behavior condition versus the socially motivated model

5. Multimodal Feedback Study

Since a robot is not limited to human-like feedback modalities, two non-human-like modalities are proposed: Firstly, a screen to display the retrieved route knowledge to the human, motivated by a previous experiment, where 14 out of 20 subjects stated that displaying the obtained knowledge helped during human-computer interaction [3]. Secondly, a pointer mounted above the robotic head to indicate the obtained directions and refer to objects within shared visual space as an equivalent for arm- and hand-gestures.

In order to evaluate the combination of human-like and non-human-like modalities, a WOz-experiment was performed. The following feedback modalities were included: (1) emotional facial expressions of the EDDIE head, (2) verbal utterances according to the Divided Dialog strategy, (3) the pointer to indicate the obtained directions, and (4) the screen positioned below the robot's head to display the route graph the robot developed during the conversation. The modalities (1) and (2) were not varied as they were evaluated previously (see Sec. 3 and 4), only (3) and (4) served as experimental conditions. In experimental conditions where the latter modalities were not used, they were in an *idle* state to ensure that they do not merely attract the user's attention by being switched on and off during the experiment. The aim of this experiment was to identify the most efficient setup regarding feedback modalities for a human interaction partner in the context of providing directions. Efficiency was measured in terms of task completion time and number of detected misunderstandings, which were experimentally induced by the human wizard. After the participants completed all runs, they were asked to rank the four feedback modalities according to their importance.

5.1. System Description

The experimental system architecture consists of several sub-systems with interconnections realized using RTDB [14] and ROS (www.ros.org) middleware. Actions of the system are generated by two modules, one for recognition and synthesis of facial expressions according to the Social Motivation Model described in Sec 4 and one for generation of the flow of dialog acts. The latter handles the occurrence of events with timing related to the progression of the dialog including verbal utterances, pointer movements and route graphs presented on the screen. The dialog structure and system described in Sec. 3 is used. Since a robot with a female voice was perceived slightly better in a previous WOz experiment on human-robot communication in public space [31], a female synthesized voice is used.

A pointer, which can be tilted in arbitrary directions, is added to EDDIE. In order to use the pointer as additional feedback modality in the route description experiment, the pointer is tilted with each direction given in verbal utterances into the corresponding direction with the beginning of the movement coinciding with the beginning of the word specifying the direction in the synthesized speech. The pointer returns to its default position at the end of the sentence. In its *idle* state, the pointer performs small random movements at irregular intervals.

Another additional feedback modality is provided by an 8” color screen integrated at a suitable viewing position. During dialog, the maps presented to the subjects initially are shown on the screen again along with the verbal reconfirmation prompts in the *confirmation* phase of the dialog. Route graphs are displayed at the beginning of the confirmation sentences for the corresponding segments. The screen shows a neutral image in its *idle* state.

5.2. Experimental Setup

The WOz experiment was set up as within subjects design, with a total of 20 participants. The participants played the so-called ”taxi-driver game” with the robot, in which they had to explain directions according to a map with an indicated route. After the participants provided the directions to the robot, it repeated the retrieved information and provided the human interaction partner with feedback according to the particular specification. To study the quality of task-related feedback, it was explored if experimentally induced misunderstanding influences the interaction. Therefore, under certain conditions the robot pretends that it misunderstood one item of the directions by giving a wrong repetition (e.g. the user says ”left” but the robot repeats ”right”). The feedback modalities are combined to investigate the conditions as shown in Tab. 1. The condition sequences were counterbalanced to prevent learning effects. To ensure comparable conditions, the wizard used a wizarding tool that strictly followed the Divided Dialog strategy. Additionally, the tool controlled both, the pointer and the screen. The robot was wizarded to pretend that it was acting autonomously.

Table 1. Experimental Conditions

Condition	Pointer	Screen	Misunderstanding
1	Idle	Idle	Off
2	Idle	Idle	On
3	On	Idle	Off
4	On	Idle	On
5	Idle	On	Off
6	Idle	On	On
7	On	On	Off
8	On	On	On

The experiment is designed to exhibit the most effective and satisfying feedback modalities to support users in direction requests in human-robot communication in public spaces. The main hypothesis is that the interaction efficiency increases with the number of employed feedback modalities. Interaction time is used as a measure for interaction efficiency.

5.3. Results

The 20 participants were counterbalanced regarding gender. Mean age was 27 years (SD 5.848). In Tab. 2 the mean task duration for runs without induced misunderstanding is shown.

Table 2. Mean Task Completion Time for Runs without Misunderstanding

Condition	Mean Task Completion Time	SD
1	00:03:29	00:00:58,79
3	00:03:11	00:00:49,63
5	00:03:11	00:00:39,11
7	00:03:10	00:00:42,52

According to the descriptive results, the dialogs in the condition without additional feedback modalities provided by the robot are less efficient regarding task completion time compared to those conditions, in which the robot provides additional feedback with either pointer, screen, or both. A trend regarding within-subjects contrasts between the condition with no additional feedback modalities and the three conditions with additional feedback modalities can be observed in a repeated measures ANOVA: $F(1,19) = 4.06$, $p = .069$. The hypothesis cannot be verified, but the trend shows that it may be verified in a similar experiment with a larger sample.

A similar analysis for runs with misunderstanding would in this particular case not make sense since not all instances of misunderstandings that the robot induced were detected from the participants and some runs needed more than one approach to clarify the directions. Thus, task completion times are less comparable in conditions with an induced instance of misunderstanding.

After the participants had performed all eight runs, they were asked to rank the four feedback modalities according to their importance, see Tab. 3. Apart from 4 participants who prefer the screen over verbal feedback, 16 participants stated that they favor verbal utterances as feedback modality. All modalities were ranked on second place by some participants, with a slight tendency towards the screen to be the second most useful modality. The feedback via facial expression was ranked in third place by most participants. 15 participants agreed on the pointer being the least useful feedback modality. The experiment shows that in a communicative scenario like this verbal feedback is the most important feedback modality. However, the screen proved to be useful in providing an additional possibility of reassurance to the participants, whereas facial expressions of the robot may make the conversation more pleasant.

Table 3. Participants' Ranking of the Feedback Modalities

Modality/Rank	1st	2nd	3rd	4th
Verbal	**16**	3	1	0
Facial Expression	0	5	**12**	3
Pointer	0	4	1	**15**
Screen	4	**8**	6	2

The participants were asked to comment on the setup and make suggestions for improvements: 12 participants stated that they liked the facial expressions, whereas 5 participants did not like it that much (”too monotonous”, ”sweet but scary”, ”makes me nervous”). One participant preferred a robot without facial expressions and one liked the facial expressions, but was not able to understand their meaning. For 8 participants the pointer was irritating or distracting. One participant assumed that it was necessary for the robot to work, whereas 2 participants stated that they would like to be able to make sense of the pointer.

In order to investigate whether the ratio of responses are different across the groups, the detected misunderstandings were analyzed using Cochran's Q test, see Tab. 4. Two participants had to be excluded from the analysis as they confused left with right and vice versa, so the induced misunderstandings could no longer be clearly assigned to the robot. We found a significant difference between the detected misunderstandings of the robot between the different feedback modalities (χ^2 (3) = 9.158, p = .027). A pairwise comparison revealed that significantly less participants detected the robot's misunderstandings in condition 8, where the robot gave feedback with pointer and screen (p = .032, ϕ = -.389), as compared to condition 6 in which it only used the screen.

Table 4. Comparison of Detected Misunderstandings from a Total of 18 Participants

Condition	2	4	6	8
# detected misunderstandings	11	12	14	6

5.4. Discussion

Contrary to our expectations, feedback variations do not impact efficiency as differences are not significant. Reasons could be small sample size or lab setting as direction retrieval dialogs are clearly oriented towards a real world setting which is highly dynamic, where actual grounding processes take place and for example gestures may be used to reference to a certain point in the landscape and not just to refer to a theoretical route provided on a map. An influence of the robot's feedback setup on the participants' subjective rating of the robot is noted based ratings of modalities and suggestions for improvements. In a communication scenario, verbal feedback is the most important component to make dialogs of this kind work. However, additional feedback modalities provide the participants with increased comfort and reassurance. A facial expression is likely to increase empathy, whereas the depiction of what the robot understood provided on a screen helps the participants to detect if the robot understood them.

6. Conclusions

An overview of pre-evaluated human-like single modalities for robots that are autonomously carrying out tasks with incomplete task knowledge in populated environments is given for approach behavior, natural language, and facial expressions. The human-like modalities of verbal utterances and facial expressions are combined with two non-human-like modalities in order to be evaluated regarding their efficiency in a multimodal feedback study to explore their eligibility for proactive information retrieval from task-related HRI.

First results indicate that interactions with only human-like modalities (verbal utterances and facial expressions) are less efficient regarding task completion time than those interactions additionally using non-human-like modalities (screen and/or pointer). However, these results have to be verified regarding their significance by a larger sample of subjects in follow-up experiments. The total ranking of all modalities reveals clear preferences for the human-like verbal utterances. This, again, confirms natural language to be the modality of choice for information retrieval in this interaction domain. Although non-human-like, displaying the retrieved route knowledge of the robot on a screen turns out to be a valuable means of reassurance for users, and thus is rated better than the modality of human-like facial expressions which seem to create more social influence on HRI. Hence, the screen seems to provide a ”bridge of reassurance” between verbal utterances which transport the exchanged information between a human and a robot, and facial expressions, which create the motivating social background for this process. The pointer, as an additional non-human-like modality, turned out to be the worst rated modality and was not readable by most of the users.

The fact that induced misunderstandings were detected significantly less by participants provided with pointer and screen feedback in parallel than those provided only with feedback via the screen in addition to human-like modalities, underlines again the positive applicability of the screen. Regarding the pointer as additional non-human-like modality it has to be explored in future experiments if it either distracts the attentiveness of other modalities (especially since the pointer is mounted above the head), or if it is a general issue that the presentation of two non-human-like modalities along with human-like modalities leads to a cognitive overload of human interaction partners and thus leads to a decrease of attention which may disturb information retrieval in terms of undetected misunderstandings.

For the IURO project, this work is a first step towards an integration of multimodality with the aim of making efficient use of both human-like, and non-human-like modalities in order to optimize information retrieval from task-related HRI.

Acknowledgements

The authors would like to thank Daniel Carton, Roderick de Nijs, and Nikos Mitsou for their substantial support in the project. This work is supported in part by the EU FP7 STREP project ”IURO - Interactive Urban Robot),” contract number 248317, see www.iuro-project.eu, within the DFG excellence initiative research cluster ”Cognition for Technical Systems - CoTeSys,” see www.cotesys.org, and by the Institute for Advanced Study (IAS), Technische Universität München, see also www.tum-ias.de.

References:

[1] 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.

[2] C. Bartneck, D. Kulic, and E. Croft, “Measuring the anthropomorphism, animacy, likeability, perceived intelligence, and perceived savety of robots.”, In Workshop on Metrics for Human-Robot Interaction, 37– 44, 2008.

[3] A. Bauer, B. Gonsior, D. Wollherr, and M. Buss, “Dialog System for Human-Robot Interaction for Asking for Directions”, Computational Linguistics, 2010, *submitted*.

[4] N. Bischof, “Untersuchungen zur Systemanalyse der Sozialen Motivation IV: Die Spielarten des Lächelns und das Problem der motivationalen Sollwertanpassung”, Zeitsch. Psychol., 204:1–40, 1996.

[5] I. Borutta, S. Sosnowski, K. Kühnlenz, M. Zehetleitner, and N. Bischof, “Generating Artificial Smile Variations Based on a Psychological System-Theoretic Approach”, In IEEE Int. Symp. on Robot and Human Interactive Communication, 2009.

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

[7] M. Buss, D. Carton, B. Gonsior, K. Kühnlenz, C. Landsiedel, N. Mitsou, R. d. Nijs, J. Zlotowski, S. Sosnowski, E. Strasser, M. Tscheligi, A. Weiss, and D. Wollherr, “Towards Proactive Human-Robot Interaction in Human Environments”, In Int. Conf. on Cognitive Infocommunications, 2011.

[8] A. Cesta, G. Cortellessa, V. Giuliani, F. Pecora, R. Rasconi, M. Scopelliti, and L. Tiberio, “Proactive assistive technology: An empirical study”, Human-Computer Interact., 255–268, 2007.

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

[10] K. Dautenhahn, M. Walters, S. Woods, K. Koay, C. Nehaniv, A. Sisbot, R. Alami, and T. Siméon, “How may I serve you?: a robot companion approaching a seated person in a helping context”, In ACM Conf. on Human-Robot Interaction, 172–179, 2006.

[11] B. Dumas, D. Lalanne, and S. Oviatt, “Multimodal interfaces: a survey of principles, models and frameworks”, Human Machine Interaction, 3–26, 2009.

[12] G. Ferri, A. Mondini, A. Manzi, B. Mazzolai, C. Laschi, V. Mattoli, M. Reggente, T. Stoyanov, A. J. Lilienthal, M. Lettere, and P. Dario, “DustCart, a Mobile Robot for Urban Environments: Experiments of Pollution Monitoring and Mapping during Autonomous Navigation in Urban Scenarios”, In ICRA Workshop on Networked and Mobile Robot Olfaction in Natural, Dynamic Environments, 2010.

[13] M. Gabsdil, “Clarification in spoken dialogue systems:”, In AAAI Symp. Workshop on Natural Language Generation in Spoken and Written Dialogue, 2003.

[14] M. Goebl and G. Färber, “A real-time-capable hard- and software architecture for joint image and knowledge processing in cognitive automobiles”, Intelligent Vehicles Symp., 737–740, 2007.

[15] B. Gonsior, S. Sosnowski, C. Mayer, J. Blume, B. Radig, D. Wollherr, K. Kühnlenz, “Improving Aspects of Empathy and Subjective Performance for HRI through Mirroring Facial Expressions”, In IEEE ISymp. Robot & Human Interactive Comm., 60–356, 2011.

[16] B. Gonsior, C. Landsiedel, A. Glaser, D. Wollherr, and M. Buss, “Dialog Strategies for Handling Miscommunication in Task-Related HRI”, In IEEE ISymp. Robot& Human Int. Comm., 2011.

[17] B. Gonsior, D. Wollherr, and M. Buss, “Towards a Dialog Strategy for Handling Miscommunication in Human-Robot Dialog”, In IEEE ISymp. Robot & Human Interactive Comm., 284–289, 2010.

[18] A. L. Gorin, G. Riccardi, and J. Wright, “How May I Help You?”, Speech Communication, 23:113–127, 1997.

[19] A. A. Grandey, G. M. Fisk, A. S. Mattila, K. J. Jansen, and L. A. Sideman, “Is service with a smile enough? Authenticity of positive displays during service encounters”, Organizational Behavior and Human Decision Processes, 96(1):38–55, 2005.

[20] H. Gross, H. Bühme, C. Schröter, S. Müller, A. König, C. Martin, M. Merten, and A. Bley, “Shopbot: Progress in developing an interactive mobile shopping assistant for everyday use”, In IEEE Int. Conf. on Systems, Man and Cybernetics, 2008.

[21] E. Hall, “Proxemics”, Current anthropology: a world journal of the sciences of man, 1968.

[22] E. Hall, ”The hidden dimension”, vol. 6, Doubleday NY, 1966.

[23] M. Heerink, B. Krose, V. Evers, and B. Wielinga, “Measuring acceptance of an assistive social robot: a suggested toolkit”, In IEEE ISymp. Robot & Human Interactive Comm., 528–533, 2009.

[24] 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.

[25] J.F. Kelley, “An empirical methodology for writing user-friendly natural language computer applications”, In ACM SIG-CHI 83 Human Factors in Computing systems, 193–196, 1983.

[26] K. Koay, E. Sisbot, D. Syrdal, M. Walters, K. Dautenhahn, and R. Alami, “Exploratory study of a robot approaching a person in the context of handing over an object”, In AAAI-SS Multi-disciplinary Collaboration for Socially Assistive Robotics, 18–24, 2007.

[27] V. Kulyukin, “On natural language dialogue with assistive robots”, In ACM Conf. on Human-Robot Interaction, 164–171, 2006.

[28] 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.

[29] C. Mayer, S. Sosnowski, B. Radig, K. Kühnlenz, ”Towards robotic facial mimicry: system development and evaluation”, In Int Symp on Robot and Human Interactive Communication, 2010.

[30] M. Michalowski, S. Sabanovic, C. DiSalvo, D. B. Font, L. Hiatt, N. Melchior, and R. Simmons, “Socially Distributed Perception: GRACE plays Social Tag at AAAI 2005”, Autonomous Robots, 22(4):385–397, 2007.

[31] N. Mirnig, S. Riegler, A. Weiss, and M. Tscheligi, “A Case Study on the Effect of Feedback on Itinerary Requests in Human-Robot Interaction”, In IEEE Int. Symp. on Robot and Human Interactive Communication, 343–349, 2011.

[32] N. Mirnig, A. Weiss, and M. Tscheligi, “A communication structure for human-robot itinerary requests”, In Int. Conf. on Human-Robot Interaction, 2011.

[33] A. Muhammad, S. Alili, M. Warnier, and R. Alami, “An Architecture Supporting Proactive Robot Companion Behavior”, In AISB Convention, 2009.

[34] R. Müller, T. Röfer, A. Lankenau, R. Musto, K. Stein, and A. Eisenkolb, “Coarse Qualitative Descriptions in Robot Navigation”, In Spatial Cognition II, 265–276, 2000.

[35] I. Nourbakhsh, C. Kunz, and T. Willeke, “The mobot museum robot installations: A five year experiment”, In IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, 2003.

[36] W. F. P. Ekman, “Felt, False and Miserable Smiles”, Journal of Nonverbal Behavior, 6:238–252, 1982.

[37] B. K. Payne, C. M. Cheng, O. Govorun, and B. D. Stewart, “An inkblot for attitudes: Affect misattribution as implicit measurement”, J. Personality and Social Psych., 89(3):277–293, 2005.

[38] M. Salem, K. Rohlfing, S. Kopp, and F. Joublin, “A Friendly Gesture: Investigating the Effect of Multimodal Robot Behavior in Human-Robot Interaction”, In IEEE Int. Symp. on Robot and Human Interactive Communication, 247–252, 2011.

[39] A. Sanfeliu and J. Andrade-Cetto, “Ubiquitous networking robotics in urban settings”, In IEEE/RSJ IROS Workshop on Network Robot Systems, 14–18, 2006.

[40] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita, “How to approach humans?: strategies for social robots to initiate interaction”, In ACM/IEEE Human-Rob. Interact., 109–116, 2009.

[41] E. Schegloff, “On the organization of sequences as a source of ”coherence” in talk-in-interaction”, Conversational organization and its development, 38:51–77, 1990.

[42] A. Schmid, O. Weede, and H. Worn, “Proactive robot task selection given a human intention estimate”, In IEEE Int. Symp. on Robot and Human interactive Communication, 726–731, 2007.

[43] O. Schrempf, U. Hanebeck, A. Schmid, and H. Worn, “A novel approach to proactive human-robot cooperation”, In IEEE Int. WS Robot & Human Interactive Communication, 555–560, 2005.

[44] H. Shi and T. Tenbrink, “Telling Rolland where to go: HRI dialogues on route navigation”, In WS on Spatial Language and Dialogue, 23–25, 2005.

[45] G. Skantze, ”Error Handling in Spoken Dialogue Systems - Managing Uncertainty, Grounding and Miscommunication”, Doctoral thesis, KTH Computer Science and Communication, 2007.

[46] S. Sosnowski, A. Bittermann, K. Kühnlenz, and M. Buss, “Design and Evaluation of Emotion-Display EDDIE”, In Int. Conf. on Intelligent Robots and Systems, 2006.

[47] R. Stiefelhagen, C. Fugen, R. Gieselmann, H. Holzapfel, K. Nickel, and A. Waibel, “Natural human-robot interaction using speech, head pose and gestures”, In IEEE/RSJ Int. Conf. Intelligent Robots and Systems, vol. 3, 2422–2427, 2004.

[48] T. Tenbrink, S. Hui, and K. Fischer, “Route instruction dialogues with a robotic wheelchair”, In Workshop on the Semantics and Pragmatics of Dialogue, 163, 2006.

[49] S. Thrun, M. Bennewitz, W. Burgard, A. Cremers, F. Dellaert, D. Fox, D. Hahnel, C. Rosenberg, N. Roy, J. Schulte, et al., “MINERVA: A second-generation museum tour-guide robot”, In IEEE Int. Conf. on Robotics and Automation, 1999–2005, 1999.

[50] M. L. Walters, K. Dautenhahn, S. N. Woods, and K. L. Koay, “Robotic etiquette: results from user studies involving a fetch and carry task”, In ACM/IEEE Int. Conf. on Human-Robot Interaction, 317–324, 2007.

[51] M. Walters, K. Dautenhahn, S. Woods, K. Koay, R. Te Boekhorst, and D. Lee, “Exploratory studies on social spaces between humans and a mechanical-looking robot”, Connection Sc., 18(4):429, 2006.

[52] 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 ACM/IEEE Int. Conf. on Human-Robot Interaction, 23–30, 2010.

[53] S. Werner, B. Krieg-Brückner, T. Herrmann, “Modelling Navigational Knowledge by Route Graphs”, In Spatial cognition II, 295–316, 2000.

[54] S. Woods, M. Walters, K. Koay, and K. Dautenhahn, “Comparing human robot interaction scenarios using live and video based methods: towards a novel methodological approach”, In IEEE Int. Workshop on Advanced Motion Control, 750–755, 2006.

[55] S. Woods, M. Walters, K. Koay, and K. Dautenhahn, “Methodological issues in HRI: A comparison of live and video-based methods in robot to human approach direction trials”, In IEEE Int. Symp. on Robot and Human Interactive Communication, 51–58, 2006.

[56] D. Wunderlich, “Wie analysiert man Gespräche? Beispiel: Wegauskünfte”, Linguistische Berichte, 53:41–76, 1978.

[57] J. Złotowski, A. Weiss, M. Tscheligi, “Navigating in public space: Participants evaluation of a robots approach behavior.”, In ACM/IEEE Int. Conf. on Human-Robot Interaction, 2012.

Name:
Barbara Gonsior

Affiliation:
Research Assistant, Institute of Automatic Control Engineering, Technische Universität München

Address:
80290 Munich, Germany

Brief Biographical History:
2009 Master of Arts (M.A.) in Psycholinguistics & Psychology, Ludwig-Maximilians-Universität München
2009 - Research Assistant on Human-Robot Interaction, Technische Universität München

Main Works:
• B. Gonsior, S. Sosnowski, C. Mayer, J. Blume, B. Radig, D. Wollherr, K. Kühnlenz, Improving Aspects of Empathy and Subjective Performance for HRI through Mirroring Facial Expressions, IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN), Atlanta, GA, USA, 2011

Membership in Learned Societies:
• IEEE - Institute of Electrical and Electronics Engineers

Name:
Christian Landsiedel

Affiliation:
Research Assistant, Institute of Automatic Control Engineering, Technische Universität München

Address:
80290 Munich, Germany

Brief Biographical History:
2011 - Visiting Student, KTH Stockholm
2011 - Dipl.- Ing. in Electrical Engineering, Technische Universität München

Main Works:
• C. Landsiedel, J. Edlund, F. Eyben, D. Neiberg, B. Schuller, Syllabification of Conversational Speech Using Bidirectional Long-Short-Term Memory Neural Networks, Proc. of the IEEE Int. Conf. on Acoustics Speech and Signal Processing (ICASSP) 2011, Prague, Czech Republic, 2011

Membership in Learned Societies:
• IEEE - Institute of Electrical and Electronics Engineers

Name:
Nicole Mirnig

Affiliation:
PhD student, ICT&S Center, University of Salzburg

Address:
Sigmund-Haffner Gasse 18, 5020 Salzburg, Austria

Brief Biographical History:
2011 - Master in communication studies, ICT&S Center University of Salzburg
2011 - PhD student on Human-Robot Interaction

Main Works:
• Mirnig, N., Riegler, S., Weiss, A., and Tscheligi, M. A case study on the effect of feedback on itinerary requests in human-robot interaction. Proc. of the 20th IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN), 2011.

Membership in Learned Societies:
• ACM - Association for Computing Machinery

Name:
Stefan Sosnowski

Affiliation:
Research Assistant, Institute of Automatic Control Engineering, Technische Universität München

Address:
80290 Munich, Germany

Brief Biographical History:
2002 - Technische Universität München
2007 Visiting Student, Massachusetts Institute of Technology
2007 - Technische Universität München

Main Works:
• K. Kühnlenz, S. Sosnowski, M. Buss, The Impact of Animal-like Features on Emotion Expression of Robot Head EDDIE, Advanced Robotics, 24(8-9), 2010.

Membership in Learned Societies:
• IEEE Robotics and Automation Society

Name:
Ewald Strasser

Affiliation:
PhD student, ICT&S Center, University of Salzburg

Address:
Sigmund-Haffner Gasse 18, 5020 Salzburg, Austria

Brief Biographical History:
2010 - Master in psychology, University of Vienna
2010 - joined ICT&S Center as PhD student on Human-Robot Interaction

Main Works:
• Strasser, E., Weiss, A., and Tscheligi, M. Affect misattribution procedure: An implicit technique to measure user experience in HRI. Proc. of the 7th ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI), New York, NY, USA, 2012.

Name:
Jakub Złotowski

Affiliation:
PhD student, ICT&S Center, University of Salzburg

Address:
Sigmund-Haffner Gasse 18, 5020 Salzburg, Austria

Brief Biographical History:
2010 Master in Interactive Technology, University of Tampere
2010 - joined ICT&S Center as PhD student on Human-Robot Interaction

Main Works:
• Złotowski, J., Weiss, A., and Tscheligi, M. Navigating in public space: Participants evaluation of a robots approach behavior. In Proc. of the 7th ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI), New York, NY, USA, 2012.

Name:
Martin Buss

Affiliation:
Full Professor (chair), Institute of Automatic Control Engineering, Technische Universität München

Address:
80290 Munich, Germany

Brief Biographical History:
1988 Research Student, Science University of Tokyo, Japan
1994-1995 Postdoctoral Researcher, Department of Systems Engineering, Australian National University, Canberra, Australia
1995-2000 Senior Research Assistant and Lecturer, Institute of Automatic Control Engineering, Technische Universität München
2000-2003 Full Professor, Head of Control Systems Group, Deputy Director of the Institute of Energy and Automation Technology, Technical University Berlin
2003 - Full Professor (chair), Institute of Automatic Control Engineering, Technische Universität München

Main Works:
• Buss, M., Hashimoto, H., Moore, J.B., Dextrous hand grasping force optimization, IEEE Transactions on Robotics and Automation, 12(3), pp. 406–418, 1996
• A. Bauer, K. Klasing, G. Lidoris, Q. Mühlbauer, F. Rohrmüller, S. Sosnowski, T. Xu, K. Kühnlenz, D. Wollherr, M. Buss, The Autonomous City Explorer: Towards Natural Human-Robot Interaction in Urban Environments, Int. Journal of Social Robotics, 1(2), pp. 127-140, 2009.

Membership in Learned Societies:
• IEEE Control Systems Society
• IEEE Robotics and Automation Society

Name:
Kolja Kühnlenz

Affiliation:
Senior Researcher, Institute of Automatic Control Engineering, Technische Universität München

Address:
80290 Munich, Germany

Brief Biographical History:
2003 - 2007 Research Assistant, Institute of Automatic Control Engineering, Technische Universität München
2004 Visiting Researcher, University of Southern California, L.A., USA
2007 - Senior Researcher, Institute of Automatic Control Engineering, Technische Universität München
2009 - Carl von Linde Junior Fellow, Institute for Advanced Study, Technische Universität München

Main Works:
• M. Karg, K. Kühnlenz, M. Buss, Recognition of Affect Based on Gait Patterns, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 40(4), pp. 1050–1061, 2010.
• H. Wu, L. Lou, C.-C. Chen, S. Hirche, K. Kühnlenz, Cloud-based Networked Visual Servo Control, IEEE Transactions on Industrial Electronics, 2012, accepted.

Membership in Learned Societies:
• IEEE Robotics and Automation Society

Name:
Manfred Tschelegi

Affiliation:
Professor and head of the HCI & Usability Unit, ICT&S Center, University of Salzburg

Address:
Sigmund-Haffner Gasse 18, 5020 Salzburg, Austria

Brief Biographical History:
1986 - Institute for Computer Science and Business Informatics (formerly Institute for Applied Computer Science and Information Systems, formerly Institute for Statistics and Informatics), University of Vienna (Research Associate, Lecturer, Associate Professor, Head of Working Group on Human-Computer Interaction, various other responsibilities)
1996 - CURE (Center for Usability Research & Engineering), Founder and Managing Director, established as independent non-profit research institute
2004 - Professor for Human-Computer Interaction & Usability, Head of HCI & Usability Unit, Deputy Head of Department for Computer Science, University of Salzburg

Main Works:
• C. Stephanidis, G. Salvendy, D. Akoumianakis, A. Arnold, N. Bevan, D. Dardailler, P. L. Emiliani, I. Iakovidis, P. Jenkins, A. I. Karshmer, P. Korn, A. Marcus, H. J. Murphy, C. Oppermann, C. Stary, H. Tamura, M. Tscheligi, H. Ueda, G. Weber, J. Ziegler. Toward an Information Society for All: HCI Challenges and R&D Recommendations. Int. J. Hum. Comput. Interaction, 1999: 1–28.

Name:
Astrid Weiss

Affiliation:
Postdoctoral Research Fellow, ICT&S Center, University of Salzburg

Address:
Sigmund-Haffner Gasse 18, 5020 Salzburg, Austria

Brief Biographical History:
2010 - PhD in Human-Robot Interaction
2010 - Postdoctoral Research Fellow at the ICT&S Center, University of Salzburg
2011 - Guest researcher at the Intelligent Autonomous System Group, University of Amsterdam

Main Works:
• Weiss, A., Igelsböck, J., Tscheligi, M., Bauer, A., Kühnlenz, K., Wollherr, D., and Buss, M. Robots asking for directions the willingness of passers-by to support robots. Proc. of the 5th ACM/IEEE Int. Conf. on Human Robot Interaction (HRI), New York, NY, USA, 2010, nominated for Best Paper Award.

Membership in Learned Societies:
• ACM - Association for Computing Machinery
• AISB - Society for the Study of Artificial Intelligence and the Simulation of Behaviour

Name:
Dirk Wollherr

Affiliation:
Senior Researcher, Institute of Automatic Control Engineering, Technische Universität München

Address:
80290 Munich, Germany

Brief Biographical History:
2001 - 2004 Technical University Berlin
2004 Visiting Researcher, The University of Tokyo
2004 - Technische Universität München
2010 - Carl von Linde Junior Fellow, Institute for Advanced Study, Technische Universität München

Main Works:
• G. Lidoris, F. Rohrmüller, D. Wollherr, M. Buss, System Interdependence Analysis For Autonomous Robots, The International Journal of Robotics Research, 30, pp. 601–614, 2011.
• D. Althoff, J. J. Kuffner, D. Wollherr, M. Buss, Safety Assessment of Robot Trajectories for Navigation in Uncertain and Dynamic Environments, Autonomous Robots (Special Issue on Motion Safety for Robots), 2011, accepted.

Membership in Learned Societies:
• IEEE Robotics and Automation Society
• IEEE Control Systems Society