Towards Proactive Human-Robot Interaction in Human Environments

Martin Buss, Daniel Carton, Barbara Gonsior, Kolja Kuehnlenz, Christian Landsiedel, Nikos Mitsou,
Roderick de Nijs, Jakub Zlotowski, Stefan Sosnowski, Ewald Strasser, Manfred Tscheligi, Astrid Weiss, Dirk Wollherr

Abstract— **This paper investigates proactive task-related Human-Robot Interaction (HRI) in human environments. The presented approach eventually aims for multi-modality by combining speech, gesture, and emotional facial mimicry. A first step is to focus on exploring the potentials and limitations of each modality in order to enable a robot to control a dialog in terms of proactive retrieval of missing task knowledge from humans in a natural and intuitive way. In this paper, each modality is investigated separately in the context of the IURO (Interactive Urban Robot) project, where a robot asks for its way to a predefined goal location.**

I. INTRODUCTION

Nowadays, robots are increasingly emerging from highly controlled, strictly defined laboratory settings and entering human environments. Museum guides [1], [2] and shopping assistance [3] for indoor environments and surveillance robots [4] and garbage collectors [5] for outdoor scenarios are some of the applications that have emerged towards this end. As a consequence, accessible human-robot interaction schemes become all the more important and more natural, robust and interactive user interfaces are needed to enable users who are unfamiliar with robots to successfully interact with them. To this end, the use and combination of different modalities in human-computer interaction is necessary for creating a more complete interaction experience [6].

In this contribution, we discuss issues related to the operation of robots in urban, public spaces as occurring in the IURO project, where an autonomous robot is given the task of finding a pre-defined location in a city purely relying on information gathered from the interaction with passers-by and without further map knowledge. The goal of this paper is to present an overview of the aspects of human-robot interaction that, within the IURO project, have been identified as essential for robots that are autonomously carrying out tasks with incomplete task knowledge in populated environments. Three different modalities are considered: a) facial expressions as a mean of providing intuitive feedback about the internal state of the robot and social cues for an emotional involvement of the partner, b) the handling of miscommunication and the analysis of user input during natural language interaction, and c) gestural interaction as a means of fallback in situations when speech recognition performance is low or for environments that are not suitable for verbal interaction, and to complement the information transfer capability offered by the verbal channel. Finally, another aspect of importance for the design of a robot system that is to proactively start interaction with humans is the way in which the robot attracts a human's attention and how it approaches possible interaction partners in a socially acceptable manner. A framework and an initial evaluation of a trajectory-based method for approaching humans is considered.

In general, research on proactive behavior in robotics has been focused on human intention recognition in order to make autonomous decisions on what task to execute [7]–[9]. Proactivity has also been researched in the context of domestic assistive robots in [10], where 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 paper we present techniques that are oriented toward scenarios where the beneficiary of the interaction is the robot. Since there is no a priori interest of the human for communication, 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.

The paper is organized as follows: In Section II, two different approaches for facial expression selection are presented. Section III describes a natural language dialog system where focus is given on handling miscommunication and the evaluation of the user input. In Section IV, the use of gestures is discussed and in Section IV, a human approach for proactive initiation of interaction is presented. Conclusions and future work are highlighted in Section VI.

II. FACIAL EXPRESSIONS

One of the modalities apt for a natural interface for human-robot-interaction is to give users feedback via facial expressions and to visualize internal states of the robot. Facial expressions during an interaction are synthesized with the robot head EDDIE. Detailed information on design and facial expression synthesis can be found in [11] and [12]. These facial expressions are used by the IURO robot during interactions to socialize with users and provide state feedback. We propose to use socialization in this proactive approach to interest people in helping the robot, enrich the exchange of social cues during the dialog and thus keep attention of the interaction partner until the necessary information has been retrieved.

The feeling of empathy towards others is an important form of socialization in human-human interaction. The EDDIE system uses two approaches to potentially generate empathy in an interaction partner, one being an approach based on the mirror neuron system and the other an approach based on a social model, which focuses on smiling. An overview on both is given in the following subsections. Experiments and evaluations are currently being undertaken to see which of the approaches works better in interaction, or if a combination is beneficial.

A. Mirroring of Expressions

According to neuroscientific research, the mirror neuron system affects empathy felt for interaction partners [13], [14]. It seems that

---

This work is part of the EU-STREP project IURO (Interactive Urban Robot), supported by the 7th Framework Programme of the European Union, ICT Challenge 2 Cognitive Systems and Robotics, contract number 248317, see also `www.iuro-project.eu`, and supported in part within the DFG excellence initiative research cluster *Cognition for Technical Systems – CoTeSys*, see also `www.cotesys.org`. Support of the TUM Institute for Advanced Study (IAS), Technische Universität München, is hereby gratefully acknowledged, see also `www.tum-ias.de`.

Institute of Automatic Control Engineering (LSR), Technische Universität München, D-80290 Munich, Germany
email: mb@tum.de, dc@tum.de, bg@tum.de, koku@tum.de, christian.landsiedel@tum.de, nmitsou@tum.de, rsdenijs@tum.de, jakub.zlotowski@sbg.ac.at, sosnowski@tum.de, ewald.strasser@sbg.ac.at, manfred.tscheligi@sbg.ac.at, astrid.weiss@sbg.ac.at, dw@tum.de

the partner's mental state is (partially) mirrored and experienced through motor, sensory and somatic actions. It is also notable that the derivation of emotional states from facial expressions leads to neural activity in brain regions responsible for the face. Thus, a robotic system would benefit from being able to recognize the emotional state of an interaction partner and provide appropriate, possibly mirrored, feedback.

IURO provides a closed-loop interaction by combining facial expression recognition and synthesis. In this combined setup, the robot is reacting through facial expressions, with the appropriate reaction being determined from the user's facial expression extracted from camera images. The information in the user's facial expression is decoded using the facial action coding system by extracting features from the camera images. These features are processed in a model based approach, featuring the Candide-III face model [15], providing an estimation of the activation of action units. This activation is then either processed as an input for the internal state of the robot or can be mapped to the joint-space of the robotic facial display. The combination of both analysis and synthesis modules allows to have a live mirroring system of facial expressions. Experiments show that this mirroring, respectively the transformation from the human facial expression to the robotic one, is seen as displaying the same facial expression on the human and the robot [16]. This can be taken as a basis for further experiments on mirroring and the induction of empathy during an interaction, e.g. the modulation of the mirrored facial expressions.

B. Smiling Based on a Social Model

It has been shown that smiling affects the subjective impression of task performances, for example the task performance of service providers is rated higher if the provider displays friendliness through smiling. The rating is even higher if the smile appears to be authentic [17]. This means that the implementation of appropriate context-sensitive smile reactions could be beneficial in human-robot-interaction. This system-theoretic approach to artificially generate various types of psychologically plausible smiles aims at making the interaction with IURO more pleasant and possibly achieve a subjective improvement of the robot's task performance.

"The system-theoretic account of smiling [18], is based on a reduced version of the Zurich Model of Social Motivation [19] and can describe the effect of smiling 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 on 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 and distinctive smile variation.

This model is capable of generating seven different types of smiles, as described in [18]. Each type of smile is primarily based on one of the three state dimensions (security, arousal, or autonomy) and occurs in context with an external disturbance of the homeostasis.

1) Trustful smile: The trustful smile originates in the security system and is based on a compensation of security-appetence. The underlying mood is often interpreted as happiness.

2) Smile of relief: The smile of relief originates also in the context of the security system. The rapid neutralization of a previous appetency of security triggers the smile.

3) Embarrassed smile: This type of smile is again based on security. In contrast to the former smiles it originates in a security-aversion, i.e. the system encounters more security than is necessary for homeostasis.

4) Anxious and surprised smile: Both smiles are triggered by the arousal system. The anxious smile originates in an aversion of arousal, e.g. a frightening event, in contrast to the appetence-based surprised smile.

5) Superior and inferior smile: These types of smiles occur during direct regulation of autonomy (compared to indirect regulation of autonomy in the previous cases), typically during conflicts. The smiles result from autonomy-appetence and aversion when one agent dominates the conflict and the other subdues.

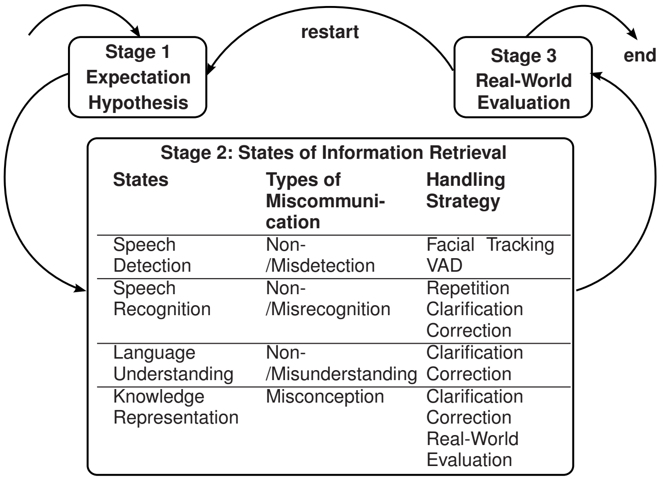
With this model, an agent is able to react to various, even unknown, situations as long as the parameters for security, arousal and autonomy can be extracted. For more detailed information on the composition of the social model, the generation of the various types of smiles and experimental results on the recognition of these smiles, please refer to [20].

III. NATURAL LANGUAGE DIALOG

Autonomous robots in populated environments are faced with knowledge gaps regarding their missing task knowledge. These should be attempted to be filled using all available information channels, including task- or environment-related information on a high level of abstraction which is available from humans. Generally, natural language is a 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 conjectures on modality selection formulated in [21], since the robot is autonomous in large portions of its behavior (navigation, action selection etc.), but also dependent 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 and the technical challenges posed by automatic speech recognition.

Robots that proactively ask humans for directions in order to extract information about their environments are still operating in very simple structured indoor environments. Coarse qualitative route descriptions can be given to a wheelchair robot [22] that navigates in an office floor. The office robot Jijo-2 [23] can learn 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 [24]. A miniature robot that can find its way in a model town by asking for directions is described in [25]. 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 proper environment modeling for situatedness of the dialog. As there is no control over the environment conditions, which may have great influence on speech recognition performance, speech recognition errors may occur frequently, and hence miscommunication has to be handled.

Several Wizard-of-Oz studies explored miscommunication and complexity arising from users giving verbal route instructions to a robot within a simulated dialogic interaction. The route instructions are carried out by the robot simultaneously during the dialog.



States	Types of Miscommunication	Handling Strategy
Speech Detection	Non-/Misdetection	Facial Tracking VAD
Speech Recognition	Non-/Misrecognition	Repetition Clarification Correction
Language Understanding	Non-/Misunderstanding	Clarification Correction
Knowledge Representation	Misconception	Clarification Correction Real-World Evaluation

Fig. 1: Dialog framework for handling miscommunication in task-related HRI

Hence, the dialog is bound to a shared perceptional space and context where the user can observe and thus directly control the actions of the robot [26], [27]. However, this paper addresses a robot that executes previously gained instructions autonomously in a real-world environment. That means complexity and the range of potential errors increase enormously within dialog due to the linguistical meta-level of talking-about-instructions which are to be carried out in the future, and unpredictable environmental impacts on speech recognition and feasibility of instructions. Within the setting envisioned in the IURO project, informational misalignment may be undetected during dialog but lead to errors during task execution based on the gained route knowledge. Thus, it is necessary to represent and evaluate the obtained task knowledge.

In order to meet the requirements mentioned above, a framework for handling miscommunication in dialogs was developed which is related to the route description task, but can be modified and extended to handle other real-world interaction tasks as well.

Originally introduced in [28], a further developed version of the framework is presented in Fig. 1. It is based on the *Theory of Perceptual Hypotheses* [29] and the *Hypotheses Theory of Social Perception* [30], which state that any (mis-)interpretation in human communication can be understood as a perceptual decision process that handles an informational stimulus stemming from the environment. This proceeds in a three-stage loop, consisting of the following stages:

1) Provision of expectation hypothesis
2) Information input
3) Confirmation (end)/disproof (restart) of the hypothesis

These consecutive stages are transferred to HRI for the context of spoken language dialog as a means of retrieving missing task knowledge and thereby handling miscommunication in human-robot interaction. The form of each stage is specified in the following.

A. Stage 1: Expectation Hypothesis

Based on a context model underlying the dialog system and any conceptual knowledge known to the robot (e.g. from previous interactions or from sensor input), a basic hypothesis on the content of verbal user input can be generated. The hypothesis usually takes into account earlier user input, if available, and anticipations deduced from dialog history. It can, for example, be specified in the form of a grammar for an user utterance or a combination of information slots to be filled.

B. Stage 2: Information Retrieval

For natural language dialog, successful information retrieval has to pass four states: *Speech Detection*, *Speech Recognition*, *Language Understanding*, and *Knowledge Representation*. At each of these states, specific types of miscommunication can occur, for which handling strategies can be deduced from human-human corpora in order to be applied during the dialog related to the corresponding state.

In order to become aware of possible miscommunication, a confidence measure indicating the extent to which the extracted information is in accordance with the hypothesis is evaluated.

C. 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 within the real world by targeted questions to confirm the desired task-status. If different and/or conflicting hypotheses show a common denominator, the task will be performed until a critical point is reached and then the evaluation cycle restarts.

IV. USING GESTURES FOR HUMAN-ROBOT NAVIGATION INSTRUCTIONS

Bauer *et al.* present in [31] a system where a robot pro-actively approaches a person and asks for directions. The directions are given through pointing gestures and then used for directed exploration. The used method deflects the body pose from the output of a stereo vision system and then estimates the pointing direction. In [32], Waldherr *et al.* introduce a hand gesture recognition based on templates. Two gestures are used to tell a robot whether it should stop or follow. By pointing the robot is shown where it can find objects on the ground that it should collect. However, the robot does not start interaction pro-actively. Breuer *et al.* [33] fit a variable hand model into the depth data of a Time-of-Flight camera to recognize gestures. Yet this approach does not work in real time, making it unapplicable for a natural interaction scenario.

The operation area of IURO is an outdoor urban environment where natural language recognition within a dialogue is a particularly complicated task. Therefore, gestures acquired by a vision-based system as shown in Fig. 2a are incorporated into the interaction scheme for several reasons. Firstly, the gesturing supports the recognition of spatial concepts that are crucial for the navigation task, such as 'left' and 'right', since these words are often combined with the respective gesture. Secondly, pointing to a principal direction leading to the searched goal can be asked for during interaction. This direction can be validated by asking multiple persons and used for navigation purposes. Lastly, the dialogue scheme explained here results in a route description containing multiple segments that are more detailed than just the principal direction. The gesture detection and interpretation approach will be explained in the following.

A. Hand Detection

For gesture detection, IURO has to perceive the hands at first. This can be achieved by a multitude of methods, such as approaches based on Haar features [34]–[36] or by combination of skin color and depth [37]. In the current implementation, we use the Microsoft Kinect[1] sensor in combination with the OpenNI framework[2], which

---

[1] `http://www.microsoft.com/kinect`

[2] `http://www.openni.org/`

provides hand tracking. In this implementation, the tracking has to be started by a waving gesture which is requested with a synthesized verbal prompt at the beginning of the interaction.

B. Detection of Pointing Direction

Estimating a pointing direction needs the definition of a vector aligned with the human arm or between the head and the pointing hand and can then provide input for directed exploration of unknown environment [31]. In our approach, we choose the hand and wrist position for this purpose. The hand position is already known from the tracking method, whereas the wrist is found by feature extraction from the detected normalized hand shape. Since depth data is provided by the Kinect sensor, a vector between the hand center and the wrist position can then be computed. Subsequently, this vector is projected to the 2D plane the robot uses for the occupancy grid that serves as a basis for the navigational path planner. The resulting 2D vector is utilized as a principal direction that determines in which direction IURO will explore its surroundings, assuming no route description is available from spoken interaction. Figure 2 shows an interaction scenario and an illustration of gesture-based navigation. Additionally, this pointing can be used to assure that an acquired route description from speech interaction principally leads into the right direction, given that the perceived pointing direction is correct.

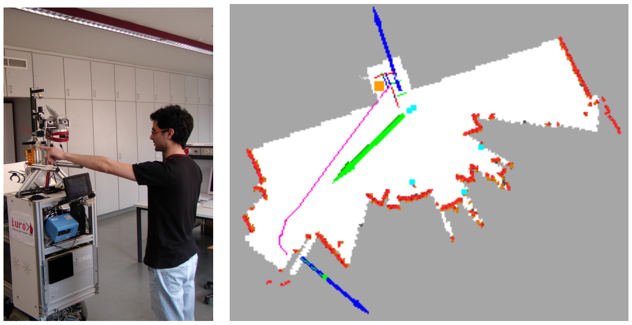

(a) (b)

Fig. 2: A person interacting with the robot and an exemplary illustration of the path planning. The green arrow indicates the perceived pointing direction, the blue arrows indicate possible exploration goals to choose, the pink line indicates the planned path to the chosen goal and the red dots represent the local obstacle map.

C. Gesture Recognition

Making use of the hand tracking, it is possible to recognize specific gestures like the pointing with a finger by incorporation of templates [32] or based on haarlets [34], as implemented in our approach. These gestures can be used to communicate more information to IURO than merely 'left' and 'right' or a single direction. As of now, the gesture recognition is only used to distinguish between gestures that yield a direction and ones that do not transport this specific information. To expand this functionality, we are evaluating the use of confirmation gestures to tell the robot that it perceived something correctly. Future plans also incorporate dynamic gestures such as waving.

V. HUMAN APPROACH

Initiating an interaction relies on positioning the robot in an appropriate position opposite the designated interaction partner. For a passive system that waits for a human to start a conversation this task can be omitted. Certainly, a robot that approaches humans and starts an interaction pro-actively should be perceived to act more natural than a passive machine.

In [38], Satake *et al.* present a method that enables a robot to approach humans pro-actively by predicting the trajectories of people in a public area. The robot then tries to catch the person's attention and initiates a conversation. [38] shows how a dextrously modeled pro-active approach strategy enhances the success of initiating human-robot-interaction.

Robovie-IV, a robotic system that roams in an office environment actively searching for interaction partners, is introduced in [39]. However, the engaging phase is passive, since the robot waits for a detected human to come close and respond to his interaction initiation. This method might constitute an alternative to the idea of going towards a person but it requires the system to be able to draw attention to itself and retain it to be successful.

Finke *et al.* describe in [40] how a person's movements and the distance to the robot can be examined to assess the interest of a person in interaction with a robot. The system waits for a human to get close and show interest rather than drawing attention and getting close by itself. Yet, the interaction initiation is pro-active because the robot decides if he should interact with the sensed person.

The IURO robot is supposed to initiate interactions proactively by both initiating the conversation and approaching the person. This forces it to plan a path to a position in front of a human in a socially acceptable way. The question arises which path or respective robot movement is most appealing to humans. [38], [41], [42] have shown preferences in robot positioning and speed as well as suitable path planning methods. However, we wanted to examine different approach trajectories, specifically taking their impression on humans into account. Moreover, we aimed to find suitable robot's approach speed and stopping distance.

In a video-based experiment conducted at the Usability Lab of the ICT&S Center (Salzburg, Austria), 30 participants were presented with pre-recorded videos showing IURO approaching a standing or walking person from front left, front right or frontal direction using the predefined trajectories shown in Fig. 3.

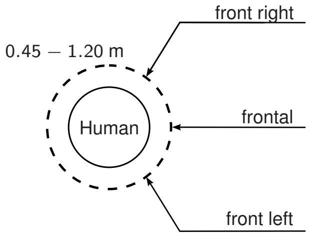


Fig. 3: Principal course of the evaluated approach trajectories for standing and walking persons.

The videos ended after IURO asked its interaction partner for directions to a specific location with a synthesized speech prompt and directions were shown by that person. The videos showed first and third person views of the approach with a male and a female actor as an interaction partner. Thus, in total six videos were acquired with each actor. Female participants were shown the videos with the female actor and male participants the video with the male actor. After each video they were asked to rate on a 9-point Likert scale how comfortable they were with the approach they had just watched.

Friedman analysis of variance by ranks was carried out to determine whether there are significant differences in participants' preferences regarding the approach trajectories. When an actor was standing still, there was a significant difference between mean rankings for the approach trajectories χ^2 (2) = 6.35, p = .04. However, no significant differences between the groups were found in conducted post-hoc tests. Nevertheless, there is a not significant trend to prefer front right over frontal approach (z = 2.32, p = .06). For approaching a walking person, no statistically significant differences were found. Yet, this could indicate both a lack of preference or the inappropriateness of the video-based study for dynamic scenes since the participants might not get the impression of walking from a video. These aspects demand further investigation as not only the approach trajectories are of interest, but also the evaluation methods. In the IURO project, a generally applicable set of methods and tools is aimed to be developed that can be applied to other robotic research projects.

The approach velocity with a defined 0.6 m/s and slowing down to 0.4 m/s at 2 m distance from the person was rated as appropriate by the participants (M = 5.09, SD = .55) on a 9-point Likert scale (1 – too slow, 5 – about right, 9 – too fast).

The last observed parameter was the stopping distance, which was defined to be within the personal space (0.45 - 1.2 m) [43] of the approached human. Participants rated this distance to be relatively comfortable (M = 4.5, SD = .79) on a 9-point Likert scale (1 – too close, 5 – about right, 9 – too far). In the dynamic condition Friedman's analysis of variance showed significant differences for the reported comfort when the robot approached from different sides (χ^2 (2) = 7.79, p = .02). However, since the pairwise comparisons were not significant, we can only report a trend for rating the stopping distance as too close in the frontal than in either front right or left approaches.

For this proactive human approach, we need to study further how appealing more complex trajectories are to humans, and how these can be integrated in a fully automated path planning scheme. Another aspect not yet taken into consideration is the selection of the person that IURO should interact with, which entails measuring attention and interest of a person towards the robot. Utilizing all of these parameters, we are then able to develop a socially acceptable approach strategy for moving and standing persons which should also feature a collision free path planning in dynamic environments.

VI. CONCLUSIONS & FUTURE WORK

In this paper, we have discussed issues specific to human-robot interaction in populated environments, where interaction partners are not necessarily expecting to be addressed by a robot and may not have previous experience with this kind of interaction. For an appropriate human-robot interaction under these circumstances, we have highlighted issues related to gestural interaction and misunderstanding handling in multimodal dialogs in a route description domain. Emotional feedback and socially inspired approach mechanisms have been addressed with respect to their importance for attracting and maintaining the interest and willingness to interact of human interaction partners.

Future work will comprise the fusion of all modalities in a combined interaction scenario. An appropriate framework for such a fusion is needed to make use of the information from all modalities to maximize information gain. For instance, when approaching a human, emotion, speech and robot movement should be evaluated if they affect the approach acceptance positively and help keeping the attention of a person on the robot. Furthermore, the potential of mutual perceptive corrections across all modalities should be considered such that e.g. speech recognition errors can be detected and corrected with the help of information gathered from gestures.

REFERENCES

[1] 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 *Proc of IEEE international conference on robotics and automation*, 1999, pp. 1999–2005.

[2] I. Nourbakhsh, C. Kunz, and T. Willeke, "The mobot museum robot installations: A five year experiment," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2003.

[3] 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 *Proc. IEEE Internat. Conf. on Systems, Man and Cybernetics (IEEE-SMC)*, 2008.

[4] A. Sanfeliu and J. Andrade-Cetto, "Ubiquitous networking robotics in urban settings," in *Proceedings of the IEEE/RSJ IROS Workshop on Network Robot Systems*, 2006, pp. 14–18.

[5] 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 *Proceedings of ICRA Workshop on Networked and Mobile Robot Olfaction in Natural, Dynamic Environments*, 2010.

[6] B. Dumas, D. Lalanne, and S. Oviatt, "Multimodal interfaces: a survey of principles, models and frameworks," *Human Machine Interaction*, pp. 3–26, 2009.

[7] O. Schrempf, U. Hanebeck, A. Schmid, and H. Worn, "A novel approach to proactive human-robot cooperation," in *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on.* IEEE, 2005, pp. 555–560.

[8] A. Schmid, O. Weede, and H. Worn, "Proactive robot task selection given a human intention estimate," in *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on.* IEEE, 2007, pp. 726–731.

[9] A. Muhammad, S. ALILI, M. WARNIER, and R. ALAMI12, "An architecture supporting proactive robot companion behavior."

[10] A. Cesta, G. Cortellessa, V. Giuliani, F. Pecora, R. Rasconi, M. Scopelliti, and L. Tiberio, "Proactive assistive technology: An empirical study," *Human-Computer Interaction–INTERACT 2007*, pp. 255–268, 2010.

[11] S. Sosnowski, K. Kuhnlenz, and M. Buss, "EDDIE - An Emotion-Display with Dynamic Intuitive Expressions," in *The 15th IEEE International Symposium on Robot and Human Interactive Communication. ROMAN 2006*, University of Hertfordshire, Hatfield, United Kingdom, 6-8 September 2006, pp. 569–574.

[12] K. Kühnlenz, S. Sosnowski, and M. Buss, "The impact of animal-like features on emotion expression of robot head eddie," *Advanced Robotics*, vol. 24, no. 8-9, 2010.

[13] 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*, 2005.

[14] V. Gallese, "The 'Shared Manifold' Hypothesis," *Journal of Consciousness Studies*, vol. 8, 2001.

[15] S. Sosnowski, C. Mayer, B. Radig, and K. Kühnlenz, "Mirror my Emotions! Combining Facial Expression Analysis and Synthesis on a Robot," in *The Thirty Sixth Annual Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB2010)*, 2010.

[16] C. Mayer, S. Sosnowski, B. Radig, and K. Kühnlenz, "Towards robotic facial mimicry: system development and evaluation," in *19th IEEE International Symposium inRobot and Human Interactive Communication*, 2010.

[17] 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*, vol. 96, no. 1, pp. 38 – 55, 2005.

[18] N. Bischof, "Untersuchungen zur Systemanalyse der Sozialen Motivation IV: Die Spielarten des Lächelns und das Problem der motivationalen Sollwertanpassung," *Zeitschrift für Psychologie*, vol. 204, pp. 1–40, 1996.

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

- [20] I. Borutta, S. Sosnowski, K. Kühnlenz, M. Zehetleitner, and N. Bischof, "Generating Artificial Smile Variations Based on a Psychological System-Theoretic Approach," in *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man)*, Toyama, Japan, 2009.
- [21] V. Kulyukin, "On natural language dialogue with assistive robots," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM, 2006, pp. 164–171.
- [22] R. Müller, T. Röfer, A. Lankenau, R. Musto, K. Stein, and A. Eisenkolb, "Coarse qualitative descriptions in robot navigation," in *Spatial Cognition II*, 2000, pp. 265–276.
- [23] 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*, vol. 19, no. 5, pp. 46–55, 2001.
- [24] 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*, vol. 22, no. 4, pp. 385–397, 2007.
- [25] 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.
- [26] G. Skantze, "Exploring human error recovery strategies: Implications for spoken dialogue systems," *Speech Communication*, vol. 45, no. 3, pp. 325 – 341, 2005, Special Issue on Error Handling in Spoken Dialogue Systems.
- [27] T. Koulouri and S. Lauria, "Exploring miscommunication and collaborative behaviour in human-robot interaction," in *SIGDIAL '09: Proceedings of the SIGDIAL 2009 Conference*. Morristown, NJ, USA: Association for Computational Linguistics, 2009, pp. 111–119.
- [28] B. Gonsior, D. Wollherr, and M. Buss, "Towards a dialog strategy for handling miscommunication in human-robot dialog," in *IEEE International Symposium on Robot and Human Interactive Communication*, 2010, pp. 284–289.
- [29] J. Bruner and L. Postman, "On the perception of incongruity: a paradigm." *Journal of Personality*, vol. 18, pp. 206–223, 1949.
- [30] W. Lilli and D. Frey, *Theorien der Sozialpsychologie [Theories of Social Psychology]*, 2nd ed. Bern: Hans Huber Verlag, 1993, vol. 1: Cognitive Theories, ch. Die Hypothesentheorie der sozialen Wahrnehmung [The Hypotheses Theory of Social Perception], pp. 49–78.
- [31] 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," in *International Journal of Social Robotics*, vol. no. 2. IEEE, 2009, pp. 127–140.
- [32] S. Waldherr, S. Thrun, R. Romero, and D. Margaritis, "Template-Based Recognition of Pose and Motion Gestures On a Mobile Robot," in *Proceedings of the fifteenth national/tenth conference on Artificial intelligence/Innovative Applications of Artificial Intelligence*. IEEE, 1998, pp. 997–982.
- [33] S. M. P. Breuer, C. Eckes, "Hand gesture recognition with a novel IR Time-of-Flight range camera: A pilot study," in *Computer Vision/Computer Graphics Collaboration Techniques Third International Conference, MIRAGE*, vol. 4418. Springer Berlin / Heidelberg, 2007, pp. 247–260.
- [34] M. Van den Bergh, F. Bosche, E. Koller-Meier, and L. Van Gool, "Haarlet-based hand gesture recognition for 3D interaction," in *Workshop on Applications of Computer Vision (WACV)*. IEEE, 2009.
- [35] M. Kölsch and M. Turk, "Robust hand detection," in *Automatic Face and Gesture Recognition, 2004. Proceedings. 6th IEEE International Conf. on*. IEEE, 2004, pp. 614–619.
- [36] P. Viola and M. J. Jones, "Robust real-time object detection," in *Computer Vision and Pattern Recognition, 2001. Proceedings. IEEE Computer Societey Conference on*. IEEE, 2001, p. 747.
- [37] M. Van den Bergh and L. Van Gool, "Combining RGB and ToF Cameras for Real-time 3D Hand Gesture Interaction," in *Applications of Computer Vision (WACV), 2011. Proceedings. ICRA'04. 2004 IEEE Workshop on*. IEEE, 2011.
- [38] S. Satake, T. Kanda, D. Glas, M. Imai, H. Ishiguro, and N. Hagita, "How to approach humans? - Strategies for Social Robots to Initiate Interaction," in *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*. ACM, 2009, pp. 109–116.
- [39] N. Mitsunaga, T. Miyashita, H. Ishiguro, K. Kogure, and N. Hagita, "Robovie-iv: A communication robot interacting with people daily in an office," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, vol. 6, 2006.
- [40] M. Finke, K. Koay, K. Dautenhahn, C. Nehaniv, M. Walters, and J. Saunders, "Hey, i'm over here-how can a robot attract people's attention?" in *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*. IEEE, 2005, pp. 7–12.
- [41] P. Althaus, H. Ishiguro, T. Kanda, T. Miyashita, and H. Christensen, "Navigation for human-robot interaction tasks," in *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, vol. 2. IEEE, 2004, pp. 1894–1900.
- [42] E. Sisbot, L. Marin-Urias, R. Alami, and T. Simeon, "A human aware mobile robot motion planner," *Robotics, IEEE Transactions on*, vol. 23, no. 5, pp. 874–883, 2007.
- [43] L. Walters Michael, K. Dautenhahn, R. t. Boekhorst, L. Koay Kheng, D. S. Syrdal, and C. L. Nehaniv, "An Empirical Framework for Human-Robot Proxemics," in *Proceedings New Frontiers in Human-Robot Interaction, K. Dautenhahn (Ed.), symposium at the AISB09 convention*. SSAISB, 2009, pp. 144–149.