The MuMMER project: Engaging human-robot interaction in real-world public spaces

Mary Ellen Foster[1], Rachid Alami[2], Olli Gestranius[3], Oliver Lemon[4], Marketta Niemelä[5], Jean-Marc Odobez[6], and Amit Kumar Pandey[7]

[1] University of Glasgow, Glasgow, UK `MaryEllen.Foster@glasgow.ac.uk`
[2] LAAS-CNRS, Toulouse, France `rachid.alami@laas.fr`
[3] Ideapark, Lempäala, Finland `olli.gestranius@ideapark.fi`
[4] Heriot-Watt University, Edinburgh, UK `o.lemon@hw.ac.uk`
[5] VTT Technical Research Centre of Finland, Tampere, Finland `marketta.niemela@vtt.fi`
[6] Idiap Research Institute, Martigny, Switzerland `odobez@idiap.ch`
[7] SoftBank Robotics Europe, Paris, France `akpandey@aldebaran.com`

http://mummer-project.eu/

Abstract. MuMMER (MultiModal Mall Entertainment Robot) is a four-year, EU-funded project with the overall goal of developing a humanoid robot (SoftBank Robotics' Pepper robot being the primary robot platform) with the social intelligence to interact autonomously and naturally in the dynamic environments of a public shopping mall, providing an engaging and entertaining experience to the general public. Using co-design methods, we will work together with stakeholders including customers, retailers, and business managers to develop truly engaging robot behaviours. Crucially, our robot will exhibit behaviour that is *socially appropriate* and *engaging* by combining speech-based interaction with non-verbal communication and human-aware navigation. To support this behaviour, we will develop and integrate new methods from audiovisual scene processing, social-signal processing, high-level action selection, and human-aware robot navigation. Throughout the project, the robot will be regularly deployed in Ideapark, a large public shopping mall in Finland. This position paper describes the MuMMER project: its needs, the objectives, R&D challenges and our approach. It will serve as reference for the robotics community and stakeholders about this ambitious project, demonstrating how a co-design approach can address some of the barriers and help in building follow-up projects.

1 Introduction

Developing an artificial agent capable of coexisting and interacting independently, naturally, and safely with humans in an unconstrained real-world setting has been the dream of robot developers since the very earliest days. In popular culture and science fiction, the prototypical image of a "robot" is precisely this: an artificial human that is able to engage fully in all aspects of face-to-face conversation. As modern robot hardware becomes safer, more sophisticated, and more generally available, there is a clear and significant consumer demand for robots that are able to coexist in everyday, real-world human environments in this way. For example, demand for SoftBank Robotics' Pepper robot has been so high in Japan that, since July 2015, each monthly run of 1000 units has sold out in under a minute [18]; also, the Jibo robot broke records on the crowdfunding site Indiegogo in July 2014 when it raised US$1million in less than a week [1]. This form of *socially intelligent* HRI [7] has therefore received a particular focus in recent years. However, while the hardware capabilities of such robots are increasing rapidly, the software development has not kept pace: even with the most recent technological developments, the most advanced such robots have generally supported limited, scripted interactions, often relying on a human operator to help with input processing and/or appropriate behaviour selection.

In the new European project MuMMER ("**Mu**lti**M**odal **M**all **E**ntertainment **R**obot"), we are developing a humanoid robot that is able to operate autonomously and naturally in a public shopping mall. The overall concept underlying MuMMER is that an interactive robot deployed in a public space such as a shopping mall should be *entertaining*: this will increase the acceptability of such a robot to the people it interacts with, therefore improving their overall experience in the mall. To ensure that the robot scenarios and interactions are engaging and entertaining, we will involve a wide number of potential users through continuous situated co-design studies, and will use the results of these studies to inform the design and operation of all components of the robot system.

In a number of recent studies, humanoid robots have been deployed in a range of public spaces, mainly in Japan. For example, Kanda et al. [11] carried out a five-week field trial of a robot in a shopping centre, involving 2343 interactions with 235 tracked participants. The robot was semi-autonomous: it used a human operator to carry out speech recognition, to monitor and override its behaviour selection, and to provide additional domain knowledge when needed. In post-experiment questionnaires, the robot was highly rated on most subjective measures, with many positive comments. Outside Japan, other successful locations for public robot deployment have included museums [12], city centres [3, 20], care homes [9], and airports [19].

Note, however, that the interactive behaviour of the robots in these previous deployments has been limited in various ways. For example, most of the robots used in these previous deployments either were not mobile, were remotely teleoperated, or were able to move only within an area delimited with environmental sensors such as floor sensors. Also, these robots are generally semi-autonomous, incorporating a human operator for operations such as speech recognition and overall behaviour selection. In general, recent experiments with robots in public spaces have tended to concentrate either purely on human-aware navigation (e.g. [6, 14]), or purely on natural-language interaction (e.g. [2]), rather than on the combination of the two.

2 Objectives

In MuMMER, we will develop a humanoid robot capable of performing its interactive tasks independently and naturally in a dynamic environment and with several users simultaneously. To meet this goal, we have defined the following set of objectives which will be used to guide the work across the project. These objectives encompass both state-of-the-art technical requirements, as well as our intentions to ensure that our work in MuMMER is grounded in real-world needs and applications throughout the project.

Objective 1: Development of an interaction robot for entertainment applications. The tasks of interaction and user experience design for entertainment are at the core of MuMMER. These highly challenging objectives are multi-party and multimodal at numerous levels (sensing, synthesis, conversation). We aim to achieve these objectives through the seamless integration of user state and mood monitoring, design of entertainment-oriented artificial social signals, and interaction management.

Objective 2: Fusion of interactive applications through a co-design process. Interactive applications tend to be specific to one or two simple tasks. However, in MuMMER, we aim to develop an interaction robot that blends different tasks (information providing, guidance, entertainment) within a single entity, relying on real user needs and interests. The goal is to ensure that the developed robotic application will be user-driven in terms of HRI, is socially and ethically accepted, and is also interesting to commercial end-users.

Objective 3: Situated-perception with on-board sensors. Rather than instrumenting the scene or people, in MuMMER we aim to design an autonomous robot able to operate robustly and naturally in diverse environments by relying only on onboard sensors. To address the resulting sensing limitations, we plan to design a truly interaction-aware perception approach combining perception algorithms aware of the robot gestures, and active sensing, along with speech synthesis and dialogue to make users implicitly or explicitly aware of perception uncertainties.

Objective 4: Automatic learning of engaging interaction strategies. We aim to develop robots which can interact naturally with humans in a social setting, being entertaining as well as helpful, and this requires learning from human interaction data. We therefore aim to develop new techniques for automatic learning of interaction strategies for socially appropriate, engaging, and robust robot action selection, based on uncertain information about the "social" state of the interaction (e.g., where the user(s) are, whether users are engaged in the interaction, what speech has been recognised, what information is in the common-ground, and the estimated perspectives of the different users). Such capabilities will create a step-change in the naturalness and acceptance of human interaction with robots.

Objective 5: Human-aware and situated interactive tasks and motion synthesis. We aim to investigate further, develop, and demonstrate a set of decisional capabilities that will allow such a robot to effectively "close the loop" between situated dialogue and robot behaviour. The robot will be able to estimate the visual perspective of its human interaction partners, their positions and postures, and—based on the current interaction and dialogue context—to synthesize the pertinent human-aware motion and interaction. This will be achieved based on a novel combination of perspective-taking, situation assessment, and human-aware task and motion planning in order to make a substantial step forward and produce more complex and validated dialogue and interaction-aware behaviours.

Objective 6: Development of new business opportunities. One of the most challenging goals of the MuMMER project will be the development of novel business models or opportunities based on the innovative components implemented during the project. We will investigate the roles and uses of interaction and entertainment robots considered as attraction along with other features within public spaces, as well as their use as one facet of companion robots for the consumer market.

(a) Ideapark shopping mall (b) *Pii Poo* Cultural Centre

Fig. 1: Deployment locations

3 Approach

In the preceding section, we set out the objectives of the MuMMER project in areas including technical development as well as the broader context in which this work is situated. In this section, we describe our concrete approach to meeting these objectives: beginning with a description of the selected deployment locations and discussing how scenarios will be co-designed with a range of stakeholders; next, summarising how the necessary technical advances will be achieved and integrated on a sophisticated humanoid robot; and finally, outlining our plans for real-world deployment and user acceptance studies of the integrated robot system throughout the project.

3.1 Stakeholders, scenarios, and co-design

Throughout the duration of the project, we will carry out continuous data collection, deployment, and evaluation activities in the Ideapark shopping centre in Lempäälä, Finland. Ideapark is the biggest commercial city in Scandinavia, with 7.3 million visitors in 2013. It contains almost 200 stores, restaurants and cafés within 100 000 square metres.

We are targeting two main classes of interaction scenario in MuMMER. In the first set of scenarios, the robot will be situated in the main shopping area of the mall (Figure 1a): in addition to supporting a set of helping tasks, such as providing guidance, information, or collecting customer feedback about the events going on, the robot will also engage fully with users, proposing events to attend or performing entertainment activities such as telling jokes and riddles, organizing contests, or distributing items such as balloons. In the other scenarios, the robot will be situated in the *Pii Poo* centre[8] (pictured in Figure 1b), which is designed as an accessible space providing inclusive cultural activities for all ages and abilities. In this case, the robot will provide entertainment activities to children or families, focussed on social games such as quizzes, memory games, and the like.

The details of the scenarios will be developed through a continuous co-design process, which will ensure that multiple relevant perspectives from consumer target groups as well

[8] `http://www.kulttuuripiipoo.fi/content/fi/1/20106/In+English.html`

as from commercial and other stakeholders will be identified, analysed and integrated into the design of the robot, the human-robot interaction, and the applications for future businesses. In particular, the co-design process will work with the technical development process to defined a set of relevant scenarios of increasing complexity, increasing the chances that the robot will perform according to the goals set, while being representative of future business applications identified with the stakeholders.

The co-design process will include workshops, seminars and scenario co-creation from both consumer and business perspectives. In addition, it will support more focused study methods to answer specific questions or tests, such as questionnaires, interviews, and observations. Scenarios and requirements will be co-created with the stakeholders to guide the technical development work. In addition, the success metrics for the development and assessment of human-interactive mobile robots for consumer markets will be co-created to ensure their validity with regard to the purpose, and the validity will be tested in the field trials. The work on success metrics will be supported by market analyses and market studies. Co-design engages people in the technical development, and MuMMER will work to integrate their contributions to the technology development and to achieve mutual learning among the different stakeholders. Co-design increases user interest and acceptance of such robots; this is a major research goal in the project.

3.2 Technical development

From a technical point of view, endowing a robot with the ability to support the desired interaction scenarios outlined above will require the integration of a number of state-of-the-art components on a sophisticated robot platform. The primary target robot platform for MuMMER is SoftBank Robotics' Pepper[9] robot (Figure 2). Pepper is a 1.20m tall, wheeled humanoid robot, with 17 joints for graceful movements, three omni-directional wheels to move around smoothly, and more than 12 hours of energy for non-stop activities, along with the ability to go back to the recharging station if required. It has a 3D camera to sense and recognize humans and their movements at a distance of up to three metres. For the initial stages of the MuMMER project, which focus on scenario development, data collection, and preliminary development of the individual technical components, we will use the default Pepper hardware which is currently available commercially. However, based on the particular needs of MuMMER, an adapted Pepper prototype will be developed during the first period of the project, including additional sensors if necessary and feasible; this will be the robot that is used for the final autonomous MuMMER system in the later stages of the project and for the later deployments.

The high-level architecture of the MuMMER system is presented in Figure 3. As shown in this diagram, the technical developments will fall into four main classes: audiovisual processing, social signal processing, interaction management, and interactive task and motion control. In the remainder of this section, we summarise how we will develop state-of-the-art components to address each of these tasks. We will use the existing Pepper interactive capabilities as a starting point wherever possible; however, we anticipate that supporting the MuMMER scenario in its full generality will require significant advances in all areas, as outlined below.

[9] `https://www.ald.softbankrobotics.com/en/cool-robots/pepper`

Fig. 2: Pepper robot (Images © SoftBank)

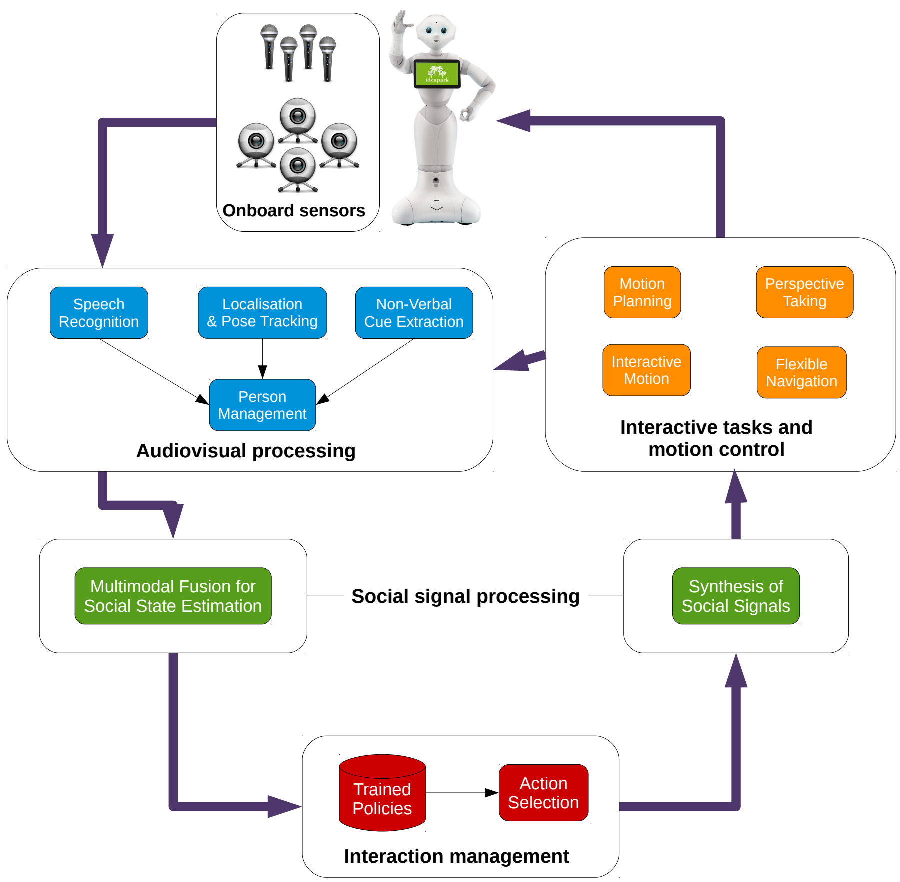

Fig. 3: Technical architecture of the MuMMER system; arrows show the main data flow

Audiovisual processing. To improve the performance of the perception modules, we will investigate several research directions. First, we will an existing real-time visual-only probabilistic multi-party head and pose tracking framework used in [16]. In particular, to handle non-stationary sensors and situations with multiple people potentially occluding each other and not facing the camera, we will leverage depth data to gain performance, e.g. by combining RGB-D based upper-body detectors (in addition to body part classifiers [17]) with very fast segmentation based tracking methods [4, 5] on the depth data stream. Second, we will develop tighter links between perception and action and will design situated algorithms that have a better knowledge of the situation and are able to exploit it. The situational context can originate from different sources (interaction state, perception state and quality, action state, etc.) and can be used in different ways: by influencing recognition and decisions, actively triggering specific perception algorithms or behaviuors (to quickly verify with a head gesture that a person that is currently outside of the field of view is still present), or via model parameter adaptation in the longer term.

Social signal processing. Social signal processing involves both the recognition of human social signals, as well as the generation and synthesis of social signals by the robot; in MuMMER, we will advance the state-of-the-art in both of these areas. For social signal recognition, we will extend algorithms for detecting and classifying user engagement and attention to apply also to the target domain of multi-party, real-world, entertaining human-robot interaction. For social signal generation, we will investigate whether "non-natural" social signals such as lights and acoustic effects can actually be socially useful in this context. We will also extend the investigation of humour from the verbal (i.e., what makes a text funny?) to the behavioural cues that can make others laugh [10] (i.e., why are two people telling the same joke not necessarily equally funny?). Finally, we will investigate how synthetic personality traits can influence the relationship between a robot and a possibly large audience (rather than a single interaction partner).

Interaction management. In this project, we will advance the state-of-the-art in interaction management in two directions. First, we will apply current state-of-the-art statistical models [13, 15, 21] to the new scenarios and tasks that arise in the context of engaging, entertaining, socially appropriate human-robot dialogue interaction. Second, we will scale up machine learning techniques to support robust human interaction with a robot in noisy, populated spaces. To support the MuMMER scenarios and tasks, the range of system actions will be extended to include communicative gestures and behaviours such as moving towards a particular person, as well as pro-active behaviour such as initiating an engagement with a person to attract their attention. In order to enable the robot to make informed decisions what to do next, given this extended range of possible system behaviours, the social state model will need to be extended with the additional relevant features. As a consequence, the extended state and action model will pose additional challenges of scalability of the machine learning techniques for state tracking and action selection that must be addressed.

Interactive tasks and motion control. Most of the prior work in the area of human-aware motion planning has mainly been from a planning perspective, with the goal of producing safe and comfortable motions. Further, the typical target scenario is not the kind of crowded space we are aiming for, with potentially many people, standing, moving, interacting and frequently blocking the path. In this context, another important dimension becomes more prominent: considering the social and cultural context of the human-robot interaction. This aspect has not been sufficiently exploited in the context of navigation planning. For example, the robot can reason from the human's perspective, analyse the feasibility that the human can provide the way for the robot, and request the human to do so through verbal or physical interaction. Similarly, the human should be able to request the same and expect a similar natural response from the robot. Such reasoning will greatly affect the operation of the state of the art navigation planning and decision making system: it requires a fusion of rich reasoning about humans, their abilities, affordances, perspective taking, interaction and navigation planning. With this, we aim to elevate the human-aware and socially-aware path planning towards more intelligent planning with social interaction possibilities (e.g., [8]).

3.3 Real-world deployment and acceptance studies

A crucial aspect of the MuMMER project plan is that the robot will be regularly deployed in Ideapark throughout the whole course of the project. For the first 3.5 years of the project, the interim deployments will be regular (i.e., every few months) but rather short-term (maximum 1–2 weeks); these visits will serve to support the co-design activities as well as to provide formative evaluation of the technical components of the system and pilot testing of the overall robot deployment process. However, at the end of the project, we will carry out a continuous, long-term user acceptance study in Ideapark lasting several months, where the robot is present in the shopping centre continuously (ideally, daily) throughout the entire period: this should allow a wide range of behaviours to be examined, and also to greatly reduce the novelty effect of the robot by providing more concrete data on how it is actually used in context across time. The results of this long-term study will serve to finalise the use cases and success metrics developed during the project, and will also provide a concrete demonstration of the utility, flexibility, and generalisability of the robot system developed in MuMMER.

As mentioned at the start of this paper, this sort of long-term, public-space deployment has only recently begun to be addressed (e.g., [9, 19]), and not very widely outside of Japan. The results of the MuMMER long-term deployment should therefore add significantly to our knowledge of how such robots are received and accepted in public spaces around Europe.

4 Current work

The project is currently in its first year, and efforts have begun towards all of the technical tasks outlines in Section 3. In particular, the first co-design workshops have been carried out with management and personnel from Ideapark and with consumers. These have resulted in an initial set of proposed scenarios, including welcoming consumers, giving information about the location of shops in the mall, playing simple quiz games and telling jokes, distributing electronic vouchers, and also posing for and sharing "selfie" images. The consumers also expressed concerns about issues such as privacy and safety, and we are taking care to address those concerns appropriately in the development of the robot system. Consultations with retailers will begin with interviews in the autumn, and the first Pepper deployment in the mall will also take place in the autumn. We have also begun to lay the groundwork for the final acceptance study by developing a questionnaire which will be administered regularly throughout the project to monitor the attitudes of Ideapark customers towards the robot: the first questionnaire will be carried out before Pepper is taken to the mall in the autumn to provide a baseline measure of these attitudes.

On the technical side, development has also begun on all of the components outlined above in Section 3, with the initial goal of supporting the scenarios and addressing the concerns arising from the first co-design studies listed above. The technical infrastructure is also being prepared to allow distributed development and testing of the components and to integration on Pepper, and all partners have also begun development and testing with the actual Pepper hardware platform. We anticipate being in a position to test an initial integrated system in the mall at the end of the project's first year (i.e., February 2017), as a step towards the eventual long-term deployment of the full system at the end of the MuMMER project's fourth year (February 2020).

5 Acknowledgements

This research has been partially funded by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 688147 (MuMMER, `mummer-project.eu`).

References

1. Annear, S.: Makers of the world's 'first family robot' just set a new crowdfunding record. Boston Daily (Jul 2014), `http://www.bostonmagazine.com/news/blog/2014/07/23/jibo-raises-1-million-six-days-record/`
2. Barker, J.: Towards human-robot speech communication in everyday environments. In: Invited Talk presented at the ICSR 2013 Workshop on Robots in Public Spaces (Oct 2013)
3. Bauer, A., Klasing, K., Lidoris, G., Mühlbauer, Q., Rohrmüller, F., Sosnowski, S., Xu, T., Kühnlenz, K., Wollherr, D., Buss, M.: The autonomous city explorer: Towards natural human-robot interaction in urban environments. International Journal of Social Robotics 1(2), 127–140 (2009)
4. Bibby, C., Reid, I.D.: Robust real-time visual tracking using pixel-wise posteriors. In: ECCV. pp. 831–844 (2008)
5. Bibby, C., Reid, I.D.: Real-time tracking of multiple occluding objects using level sets. In: CVPR. pp. 1307–1314 (2010)
6. Clodic, A., Fleury, S., Alami, R., Chatila, R., Bailly, G., Brethes, L., Cottret, M., Danes, P., Dollat, X., Elisei, F., Ferrane, I., Herrb, M., Infantes, G., Lemaire, C., Lerasle, F., Manhes, J., Marcoul, P., Menezes, P., Montreuil, V.: Rackham: An interactive robot-guide. In: Proceedings of RO-MAN 2006. pp. 502 –509 (Sep 2006)

7. Dautenhahn, K.: Socially intelligent robots: dimensions of human-robot interaction. Philosophical Transactions of the Royal Society B: Biological Sciences 362(1480), 679–704 (2007)
8. Dondrup, C., Hanheide, M.: Qualitative constraints for human-aware robot navigation using velocity costmaps. In: Proc. RO-MAN (2016)
9. Hebesberger, D., Dondrup, C., Koertner, T., Gisinger, C., Pripfl, J.: Lessons learned from the deployment of a long-term autonomous robot as companion in physical therapy for older adults with dementia: A mixed methods study. In: The Eleventh ACM/IEEE International Conference on Human Robot Interaction. pp. 27–34 (2016)
10. Hurley, M., Dennett, D., Adams Jr., R.: Inside Jokes. MIT Press (2011)
11. Kanda, T., Shiomi, M., Miyashita, Z., Ishiguro, H., Hagita, N.: A communication robot in a shopping mall. Robotics, IEEE Transactions on 26(5), 897–913 (Oct 2010)
12. Karkaletsis, V., Konstantopoulos, S., Bilidas, D., Androutsopoulos, I., Lampouras, G., Malakasiotis, P., Trahanias, P., Baltzakis, H.: Natural interaction with personality and dialogue enabled robots. In: Proc. Hellenic Artificial Intelligence Conference (SETN). Lecture Notes in Artificial Intelligence, vol. 6040. Athens, Greece (May 2010)
13. Keizer, S., Foster, M.E., Wang, Z., Lemon, O.: Machine learning for social multi-party human-robot interaction. ACM Transactions on Intelligent Interactive Systems 4(3), 14:1–14:32 (Oct 2014)
14. Kruse, T., Pandey, A.K., Alami, R., Kirsch, A.: Human-aware robot navigation: A survey. Robotics and Autonomous Systems 61(12), 1726–1743 (2013)
15. Lemon, O., Pietquin, O. (eds.): Data-driven Methods for Adaptive Spoken Dialogue Systems: Computational Learning for Conversational Interfaces. Springer (2012)
16. Sheikhi, S., Khalidov, V., Klotz, D., Wrede, B., Odobez, J.M.: Leveraging the robot dialog state for visual focus of attention recognition. In: Proc. of Int. Conference on Multimodal Interfaces (ICMI) (2013)
17. Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Kipman, A., Blake, A.: Real-time human pose recognition in parts from single depth images. In: CVPR (2011)
18. SoftBank Robotics Corp.: Press releases 2016, `http://www.softbank.jp/en/corp/group/sbr/news/press/2016/`
19. Triebel, R., Arras, K., Alami, R., Beyer, L., Breuers, S., Chatila, R., Chetouani, M., Cremers, D., Evers, V., Fiore, M., Hung, H., Ramirez, O.A.I., Joosse, M., Khambhaita, H., Kucner, T., Leibe, B., Lilienthal, A.J., Linder, T., Lohse, M., Magnusson, M., Okal, B., Palmieri, L., Rafi, U., van Rooij, M., Zhang, L.: Spencer: A socially aware service robot for passenger guidance and help in busy airports. In: Proceedings of the 10th Conference on Field and Service Robotics (FSR) (2015)
20. Weiss, A., Igelsbock, J., Tscheligi, M., Bauer, A., Kuhnlenz, K., Wollherr, D., Buss, M.: Robots asking for directions: The willingness of passers-by to support robots. In: Proceedings of HRI 2010. pp. 23 –30 (Mar 2010)
21. Yu, Y., Eshghi, A., Lemon, O.: Training an adaptive dialogue policy for interactive learning of visually grounded word meanings. In: Proc. SIGDIAL (2016)