Information Retrieval System for Human-Robot Communication – Asking for Directions

Andrea Bauer, Dirk Wollherr and Martin Buss

***Abstract*— The creation of a robot capable of navigating in unknown urban environments without the use of GPS data or prior map knowledge is envisioned in the *Autonomous City Explorer (ACE)* project. The robot has to retrieve direction information solely by interacting with humans. This work presents a human-robot communication system that enables the robot to ask for directions and store the retrieved route information as internal knowledge. The system incorporates theories from linguistics in a mixed-modalities communication interface. It stores acquired information into a topological route graph which is used to give feedback to the human and to navigate in unknown environments.**

I. INTRODUCTION

Robots are gradually moving from structured industrial settings into our daily lives. Thus, in future scenarios human-robot interaction (HRI) becomes very important. Prominent examples of interacting robots are tour guides for museums [1], [2] and shopping malls [3], that successfully relay useful pre-compiled information to humans.

Systems which are to assist humans in a flexible and versatile way will not always possess all the information required to fulfill their task. A central aspect of intelligent autonomous behavior is therefore the ability to interact in order to retrieve information. The great majority of HRI research is based on the assumption that the robot has complete knowledge of its environment. The few exceptions include [4], where a space robot asks for information in cooperative manipulation tasks, [5], in which a robot asks for the way at a robotics conference, [6], in which a miniature robot finds its way in a model town by asking for directions, and [7], where a robot creates a map of its environment by exploring it and asking a human to label areas of interest. However, all of these projects operate in extremely structured indoor environments and interaction is not initiated by the robot but by the human partner.

The *Autonomous City Explorer (ACE)* project [8] envisions a robot that can autonomously navigate to a given unknown goal location in an urban environment solely by asking passers-by for directions. The robot had no prior map knowledge or GPS data, as the research question was whether a robot can complete a task without those data that might be missing in some environments. The robot had to ask passers-by for directions on its way in order to acquire the necessary direction information, build an internal representation of that information, and follow the route.

As a principal component for achieving the goal of the project a system for asking humans for directions and building a spatial knowledge representation is created. From the direction information given by humans the robot builds a topological route graph as an internal representation of the knowledge, which is verified while following the route and transforms into a metric route graph. Classical route graphs are described in [9] and in [10] they are applied to modeling navigational knowledge.

The system used principles from linguistics on asking for directions and deictic communication, as these render the communication more natural and intuitive to humans. The human-robot communication in this work is achieved through several means. Humans can give information to the robot through gestures and input on a touch screen, while the robot uses speech, text and images, to communicate to the human. Deictic communication between human and robot has been investigated in [11], where the human points out certain objects in a room to a robot by speech and gesture.

This paper provides an overview of the route graph acquisition system. The necessary linguistics background is outlined in Sec. II. The information retrieval system is presented in Sec. III, focussing on the human-robot interaction and the knowledge representation as route graphs. Sec. IV gives an overview of the implementation of the system on the *ACE* robot. An outdoor experiment with *ACE* is presented in Sec. V. A conclusion and an outlook are provided in Sec. VI.

II. BACKGROUND FROM LINGUISTICS

Linguistics identify rules and structures of human-human communication. These findings should be taken into account when aiming at intuitive and unambiguous human-robot communication, as ”vagueness is one of the most salient, but also one of the most effective features of natural language” [12]. Two findings of linguistics are particularly relevant to asking for directions, namely a structure of asking-for-directions dialogs and the problem of aligning personal reference systems, called origines.

A structure of asking-for-directions dialogs with four phases has been identified in [13]:

1) *Introduction*: The asker addresses a respondent and defines the task, i.e. giving directions to a specified goal location.
2) *Giving directions*: The respondent provides the necessary information.
3) *Confirmation*: Either of the two partners confirms the information. In this phase further inquiries can be made.
4) *Conclusion*: The asker thanks the respondent.

Institute of Automatic Control Engineering (LSR), Technische Universität München, D-80290 Munich, Germany, {`ab, dw, mb`}`@tum.de`

For humans this schematic structure is very flexible, i.e. some phases may be interchanged or recur. Nevertheless it is a well-proven guideline from human-human communication which can be easily applied to human-robot interaction.

The biggest source of misunderstanding when giving or receiving directions is the fact that every person has a *personal reference system*, known in linguistics as the origo [14]. As outlined in Fig. 1 a point of interest (PoI) can be described as being located in opposite directions depending on the reference system O_h or O_r, rendering the direction information ambiguous.

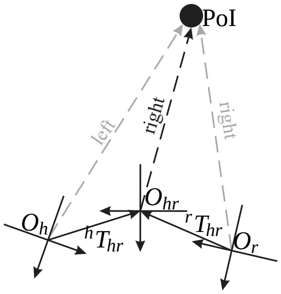


Fig. 1. A PoI is referred to differently from the personal reference systems, or origines, of two dialog partners (O_h and O_r). After coordinate transformations (${}^rT_{hr}$ and ${}^hT_{hr}$ respectively) both partners use a common reference system (O_{hr}) to refer to a PoI in an unambiguous way.

The dialog partners have to agree on a common reference system (O_{hr}) at the beginning of the *Giving Directions* phase of the dialog, to avoid ambiguity of information. When information is given relative to a common reference system, it is unambiguous.

Humans use different means of agreeing on a common reference system, the most common are gestural pointing in an initial direction or verbal pointing to a reference object. In mathematical terms, this pointing introduces a common reference system, O_{hr}, that is oriented towards the indicated direction, and specifies the necessary transformations, ${}^rT_{hr}$ and ${}^hT_{hr}$. After that both partners understand that subsequent direction information is given relative to the last information.

Both, the structure of asking for directions dialogs, and the principle of agreeing on a common reference system should be taken into account when designing a robotic system that is supposed to ask a human for directions.

III. ROUTE GRAPH ACQUISITION SYSTEM

In this work a robot has to interact with humans to retrieve missing information necessary for task completion. In specific the robot asks passers-by for directions to an unknown goal location and builds an internal representation of that information.

A route graph acquisition system retrieves spatial information from an asking-for-directions dialog and represents it as a topological route graph, that can be used for giving feedback to the human, checking the input for plausibility, and navigating. It can even be updated to a metric route graph with sensor data from the environment. An overview of the process leading from HRI to a representation of the spatial knowledge of the robot is given in Fig. 2.

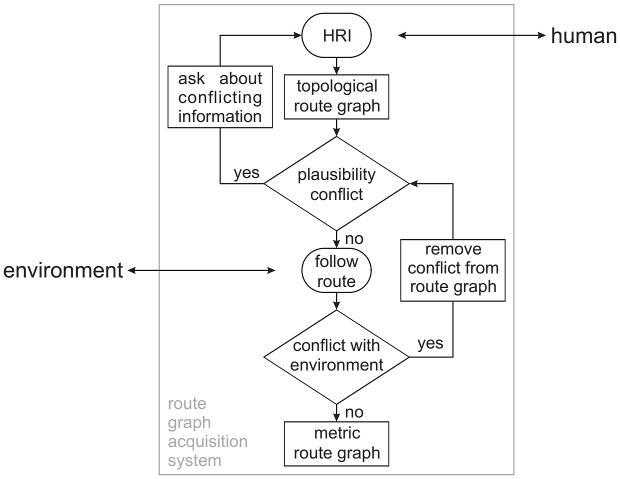


Fig. 2. Overview of the route graph acquisition system: route information given by humans is transformed into a topological route graph and checked for plausibility. The route graph is used for navigation and for giving feedback to the human partner. While following the route the robot updated the graph to a metric route graph from sensor data.

During HRI the human gives direction information to the robot, which is stored as a topological route graph. The information is checked for plausibility by comparing it to the existing route knowledge. If a plausibility conflict is found the human is asked again about the conflicting information. If the information is plausible, the robot starts following the route and collecting metric sensor data of the environment, which is used to update the graph to a metric route graph. If the sensor data from the environment conflicts with the topological route graph, this particular information is deleted from the graph and a human is asked for the correct information.

A. Human-Robot Interaction

The HRI is the sole source of information for the robot to build the topological route graph upon. Therefore it is crucial that the communication between human and robot is successful, i.e. that the robot retrieves unambiguous information. To this end, the principles from linguistics described in Sec. II are integrated in the HRI system of the robot.

The structure of asking-for-directions dialogs is adopted by the HRI system, in order to make the communication intuitive to the human, who is used to this structure. According to the *Introduction* phase of the asking-for-directions dialog the robot has to address a human and state the task of giving directions to a certain goal. After that the *giving directions* phase of the dialog begins where the human gives the necessary direction information to the robot. The *Confirmation* phase of the asking-for-directions dialog can occur after the *Giving directions* phase or simultaneously, in which case the robot has to give feedback about every new information immediately after it was given by the human, which would allow the human to correct information instantly if required.

When the human finishes giving directions the robot has to thank the human for the help and can start moving along the given route, corresponding to the *Conclusion* phase.

In order to render the route information given by the human unambiguous the *personal reference systems* of human and robot have to be aligned at the beginning of the *giving directions* phase. To achieve this, the human is asked to give the first direction by a pointing gesture. The reference system of the robot is then aligned to a *common reference system* of human and robot, that is oriented towards the direction that was pointed by the human. After that the human can give subsequent direction information relative to the common reference system.

During the dialog new route information is stored in a topological route graph by the robot and processed as described in the following.

B. The Route Graph

The information retrieved through HRI is stored as an internal representation of the path, this allows for navigation and for giving feedback to human dialog partners. The directions are first stored in a directed *topological route graph* G^T, which subsequently is updated at each intersection with the sensor data to a directed *metric route graph* G^M.

The topological route graph $G^T \langle \text{nodes, edges} \rangle$ is constructed where nodes $N_j^T(x_j^i, y_j^i, c_j)$ represent intersections and edges $E_j^i(N_i, N_j)$ denote actions connecting intersections.

A node $N_j^T(x_j^i, y_j^i, c_j)$ includes coordinates (x_j^i, y_j^i) that indicate the relative direction from the previous node N_i and a certainty value c_j that represents the accuracy of the coordinates. The certainty value ranges from 0, for an unknown direction, to 1 for a direction that has been verified by successfully reaching node N_j^T. A certainty value c_j in $(0,1)$ reflects the fact that the information was given by a human and has not yet been verified by the robot. The value is assigned such that

$$c_j = \frac{0.9(n_c^j+1)+0.1(n_r^j+1)}{n_c^j+n_r^j+2},$$

depending on the number of confirmations n_c^j and the number of rejections n_r^j of information by humans. Thus the certainty value of a node is initialized with 0.5, and is increased through a confirmation by another human or decreased by a rejection respectively.

The robot starts only with the knowledge of the start position $N_s^M = (0,0,1)$ and a given goal $N_g^T = (x_g^{g-1}, y_g^{g-1}, 0)$ with an unknown position.

When the robot retrieves new route information, a new edge $E_n^{n-1}(N_{n-1}, N_n)$ and node $N_n^T(x_n^{n-1}, y_n^{n-1}, 0.5)$ is inserted into the topological route graph.

The internal representation of the route is transformed into a METRIC ROUTE GRAPH, when the next intersection is reached by the robot. The corresponding node N_n^T in the topological route graph G^T is updated to a node N_n^M with metric coordinates from sensor information. As soon as the robot reaches its goal, the spatial knowledge is completely represented as a metric route graph G^M, which can be used later on, e.g. to find the way again, or to share the route knowledge with humans or other robots.

IV. IMPLEMENTATION

The route graph acquisition system has been implemented on the *ACE* robot. Fig. 3 shows *ACE* with principal hardware components. The robot comprises a differential wheel mobile platform, laser range finders, a stereo vision system, a touch screen, a loud speaker, and a small screen with an animated mouth. For more details on the hardware please refer to [8].

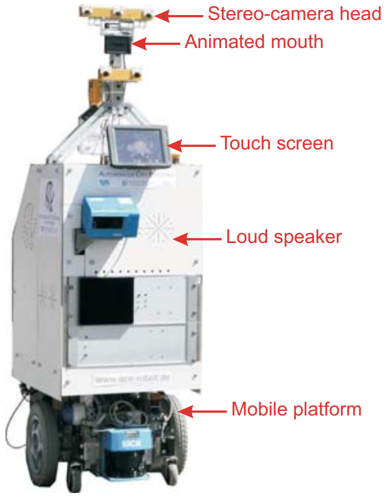


Fig. 3. The human-sized *ACE* robot with principal hardware components.

The robot relays information to the human firstly through speech over a loud speaker. The speech is synthesized using MaryTTS [15] (the human can choose between English and German). Synchronously an animated mouth is displayed (viseme-based animation) on a small monitor below the camera head. Additionally the robot can present images and text to the human on a touch screen.

As robustness to environmental disturbances such as noise is an important requirement for the system, the touch screen, is also used as the main means of input from the human.[1] Moreover the human can provide direction information via pointing gestures. For details about the gesture recognition please refer to [16].

The text the robot speaks is simultaneously displayed on the touch screen for robustness and convenience, along with informative images, e.g. the robot's view. A graphical user interface (GUI), displayed on the touch screen, comprises buttons for possible answers and buttons that allow the user to change the language, go back to the previous step in the dialog, or quit the interaction.

A finite state machine (FSM) is responsible for interfacing with the hardware for communication, navigation and vision

---

[1] Speech is not used as a input mode as speech recognition does not work robustly in outdoor environments, because traffic and human noises have the same frequency band as the signals from the human dialog partner.

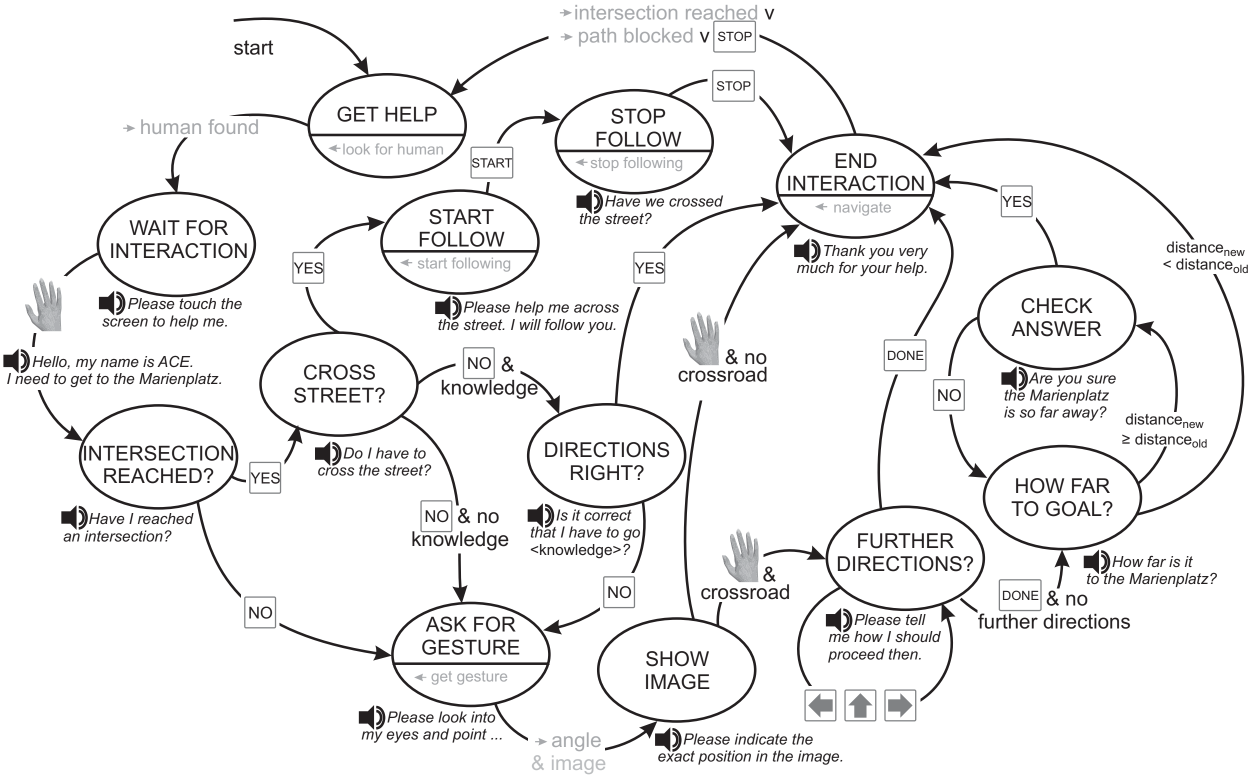
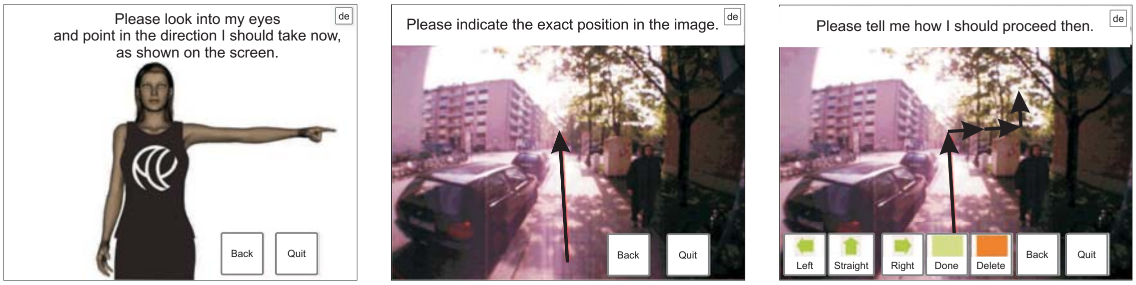


Fig. 4. Finite state machine for HRI. Transitions can be triggered by buttons that are pressed (boxed icons), when the screen is touched (hand icons), or by signals from the navigation or vision system of the robot (grey). The text the robot speaks is noted under the respective state (speaker icons).



Please look into my eyes and point in the direction I should take now, as shown on the screen.

Please indicate the exact position in the image.

Please tell me how I should proceed then.

(a) (b) (c)

Fig. 5. Example screenshots of the touch screen, where the personal reference systems (a, b) are aligned and directions are given by the human (b, c).

processing. The FSM chooses an appropriate interaction behavior depending on the situation and controls the asking-for directions dialog. Various interaction behaviors have been implemented for different situations, for example for interacting at intersections, when crossing streets, or in a pedestrian area. The FSM is depicted in Fig. 4, where the text the robot utters is noted under each respective state, marked by a speaker icon.

Transitions between the states can be triggered by inputs from humans, touch is marked by hand icons, buttons pressed are marked by boxed icons, or by signals, shown in grey, from the navigation or vision system of the robot, see [17], [16] respectively.

The states ASK FOR GESTURE and SHOW IMAGE of the FSM represent the process of aligning the personal reference system to a common reference system, screen shots of the GUI in those states are depicted in Fig. 5a and Fig. 5b respectively. State SHOW IMAGE is included as a supplement, because the gesture recognition has a significant uncertainty in cluttered outdoor environments. Fig. 5c shows the GUI in state FURTHER DIRECTIONS?, in which the human partner can provide further route information over buttons on the touch screen.

As crossing the street is a critical safety issue for an autonomous robot in an urban environment *ACE* asks a human to walk in front of it and cross the street when it is safe to do so. The robot deduces whether it has to cross an intersection from the topological route graph and the previous route, or if in doubt asks the human about it.

V. OUTDOOR EXPERIMENT

An outdoor experiment was conducted on 31 August 2008, where the *ACE* robot had to reach Marienplatz (the central square of Munich), starting at the campus of Technische Universität München. The robot did not have any prior map knowledge or GPS sensors, and therefore had to ask passers-by in order to complete its task. The robot was equipped with a route graph acquisition system for human-robot communication as introduced above. The distance *ACE* had to cover was approximately 1.5 km, partly on sidewalks in a traffic zone and partly in a crowded pedestrian area.

For a video of the experiment please refer to [18].

A. Results

The robot reached its goal after about 5 hours and interacted with 38 passers-by. The large number of people interacting arises from the fact that many of the interactions were started by curious passers-by. This also explains the relatively long duration of the experiment. In specific, the robot interacted 12 times until it reached the pedestrian area. All of these interactions were self-initiated, 6 of them at intersections. On the remaining course through the pedestrian area the robot only initiated 5 interactions and was interrupted by passers-by 21 times. The interaction partners of the robot ranged in their ages from little children to elderly people, and were of both genders. Fig. 6 shows *ACE* interacting with passers-by in the pedestrian area.

Fig. 6. The *ACE* robot interacting with passers-by in downtown Munich.

The HRI was effective in terms of retrieving route knowledge through communication. Most of the humans who gave directions to *ACE* had not interacted with a robot before and were not instructed beforehand on how to communicate with it (except for a few times when the researchers had to interact with the robot, due to an absence of other passers-by). As the interaction with uninstructed passers-by was successful, it is concluded that applying the structure of asking-for-directions dialogs to the interaction system and aligning the reference systems made the communication with the robot intuitive for humans. The process of having the human dialog partner point in the initial direction, the robot had to take, was effective for aligning the personal reference systems. There was only one situation in which the robot was sent in the wrong direction by a passer-by, but this wrong information was corrected the next time, the robot had to stop because the path was blocked.

While interacting the robot formed spatial knowledge and represented it as a route graph, an example of which is shown in Fig. 7. The graph is depicted along with a corresponding occupancy grid in a satellite image of the real environment. The robot was currently positioned at node N_2, where it had retrieved topological knowledge of the route that lied ahead (white). The direction from N_2 to N_3 was given through a gesture, further directions were given through buttons on a touch screen. The part of the route the robot has already covered had been transformed into a metric route graph (black).

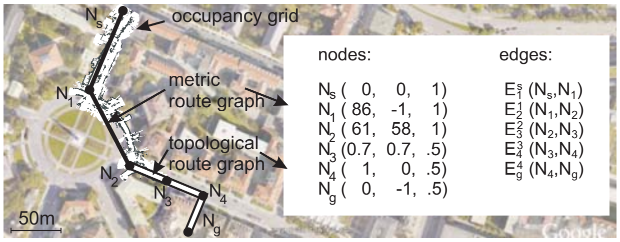


Fig. 7. Example of an occupancy grid retrieved through navigation (on the right) and a route graph gained by HRI. Shown are the metric route in [m] already covered by the robot (black) and the topological route graph given by humans (white). Right: nodes and edges.

B. Lessons Learned

As humans are used to very flexible natural human-human dialog on the one hand and rather restricted human-machine interfaces on the other hand, the communicating robot *ACE* was a novelty somewhere between those two extremes. Thus for some humans it seemed hard to assess the communicative abilities and boundaries of the robot and therefore to act accordingly. For example, many interaction partners of *ACE* expected the robot to be able to understand speech at first and tried to answer through natural language until they realized that they had to use the touch screen to communicate. Another example of a problem some humans had during the interaction with *ACE* was to make a pointing gesture that is recognizable by the robot. A similar mismatch between the users' conceptual model and the system design has been described in [19], which concludes that it is necessary to familiarize new users with the functionality of a robotic system.

A solution to the problem that humans do not know the boundaries of the communicative abilities of the robot, is to provide detailed instructions, where the communication is restricted and to provide backup solutions for the case that the communication fales. For example, to let humans understand how they could point in a way, that the robot could understand, an animation of a person pointing with an extended arm, as shown in Fig. 5, was displayed on the touch screen, which the human was asked to imitate. The problem, that humans often expected the robot to be able to understand speech, could be solved if the robot recognized the act of speaking and the human would be pointed to the fact that the robot cannot understand speech.

The design of the FSM, that included the dialog structure, known from human-human communication, and a strategy for aligning personal reference systems, was efficient in general. It could be improved in terms of asking only the relevant questions, e.g. not asking whether the robot had to cross the street, if the sensor information was more reliable.

The robot was sent in the wrong direction only once, otherwise no conflicting information occurred. However this shows that the plausibility check needs to be refined, not only asking the human about conflicting information, but ignoring the conflicting information altogether.

Crossing intersections were the most unsafe situations for the robot, as it was moving relatively slowly and in one situation it did not reach the other side of the street before the lights turned red and cars had to wait for the robot to leave the street.

VI. CONCLUSIONS AND FUTURE WORKS

A. Conclusions

In this paper an interaction system was presented that retrieves route information through human-robot communication and stores it in a route graph. Theories from linguistics are used in the design to make the communication more intuitive to humans and to avoid ambiguous information. In specific a structure for asking-for-directions dialogs and a strategy for aligning personal reference systems are included in a finite state machine that controls the dialog. The system handles the communication with passers-by, the acquisition of the relevant information from humans, plausibility checks, and the internal representation of the information as topological route graphs.

The system presented here has been used in an outdoor experiment in the city of Munich, where an autonomous robot had to find its way to a designated goal location solely on information given by humans and local sensors, as it had no GPS or map knowledge. The experiment showed that the system was effective in terms of human-robot communication and information retrieval.

B. Future Works

A major problem that occured during human-robot communication was that it was hard for humans to assess the restrictions in the communicative abilities of the robot. Therefore the system will be adapted to provide more instructions on the communication and backup solutions where the communication is restricted, whil simultaneously the abilities of the robot will be expanded to more robust gesture recognition and speech recognition. The plausibility check will be refined to handle conflicting data more reliably. Also the robot will not only wait and ask passers-by to approach it but will approach them itself in a biologically inspired way. The current system does not cooperate with GPS data, which will be included as future work.

VII. ACKNOWLEDGMENTS

This work is supported in part within the DFG excellence initiative research cluster *Cognition for Technical Systems – CoTeSys*, see also `www.cotesys.org`, and within the European Commission as part of the *Robot@CWE project*, see also `www.robot-at-cwe.eu`.

Disclaimer: The authors are solely responsible for the content of this paper, it does not necessarily represent the opinion of the European Community and the DFG.

REFERENCES

[1] W. Burgard, A. B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, “Experiences with an interactive museum tour-guide robot,” *AI*, vol. 114, pp. 3–55, 1999.

[2] S. Thrun, M. Beetz, M. Bennewitz, W. Burgard, A. Cremers, F. Dellaert, D. Fox, D. Haehnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, “Probabilistic algorithms and the interactive museum tour-guide robot minerva,” *International Journal of Robotics Research*, vol. 19 (11), pp. 972–999, 2000.

[3] H.-M. Gross, H.-J. Böhme, 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] T. W. Fong, I. Nourbakhsh, R. Ambrose, R. Simmons, A. Schultz, and J. Scholtz, “The peer-to-peer human-robot interaction project,” in *AIAA Space 2005*, 2005.

[5] M. P. Michalowski, S. Sabanovic, C. F. DiSalvo, D. B. Font, L. M. Hiatt, N. Melchior, and R. Simmons, “Socially distributed perception: GRACE plays social tag at AAAI 2005,” *Autonomous Robots*, vol. 22, pp. 385–397, 2007.

[6] 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.

[7] H.Httenrauch, K. Eklundh, and E. A.Green, “Investigating spatial relationships in human-robot interaction,” in *Proceedings of the International Conference on Intelligent Robots and Systems*, 2006.

[8] G. Lidoris, K. Klasing, A. Bauer, T. Xu, K. Kühnlenz, D. Wollherr, and M. Buss, “The autonomous city explorer project: Aims and system overview,” in *International Conference on Intelligent Robots and Systems*, 2007.

[9] B. Kuipers, “A hierarchy of qualitative representations for space,” in *International Workshop on Qualitative Reasoning about Physical Systems*, pp. 113–120, AAAI Press, 1996.

[10] S. Werner, B. Krieg-Brückner, and T. Herrmann, “Modelling navigational knowledge by route graphs,” in *Spatial Cognition II* (C. Freksa, ed.), pp. 295–316, Springer, 2000.

[11] O. Sugiyama, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita, “Natural deictic communication with humanoid robots,” in *International Conference on Intelligent Robots and Systems*, 2007.

[12] W. Klein, *Spatial Orientation: Theory, Research, and Application*, ch. Deixis and spatial orientation in route directions, pp. 283–320. New York: Plenum, 1980.

[13] D. Wunderlich, “Wie analysiert man Gespräche? Beispiel: Wegauskünfte,” *Linguistische Berichte*, vol. 58, pp. 41–76, 1978.

[14] K. Bühler, *Theory of language*. Benjamins, 1990.

[15] M. Schröder and J. Trouvain, “The german text-to-speech synthesis system mary: A tool for research, development and teaching,” *International Journal of Speech Technology*, vol. 6, pp. 365–377, 2003.

[16] Q. Mühlbauer, S. Sosnowski, T. Xu, T. Zhang, K. Kühnlenz, and M. Buss, “Navigation through urban environments by visual perception and interaction,” in *International Conference on Robotics and Automation*, 2009.

[17] G. Lidoris, F. Rohrmüller, D. Wollherr, and M. Buss, “The autonomous city explorer (ace) project - mobile robot navigation in highly populated urban environments,” in *International Conference on Robotics and Automation*, 2009.

[18] A. Bauer, K. Klasing, T. Xu, S. Sosnowski, G. Lidoris, Q.Mühlbauer, T. Zhang, F. Rohrmüller, D. Wollherr, K. Kühnlenz, and M. Buss, “The autonomous city explorer project,” in *International Conference on Robotics and Automation*, 2009.

[19] A. Green, B. Wrede, K. S. Eklundh, and S. Li, “Integrating miscommunication analysis in natural language interface design for a service robot,” in *International Conference on Intelligent Robots and Systems*, pp. 4678–4683, 2006.