Research Issues for Designing Robot Companions: BIRON as a Case Study

B. Wrede, A. Haasch, N. Hofemann, S. Hohenner, S. Hüwel,
M. Kleinehagenbrock, S. Lang, S. Li, I. Toptsis, G. A. Fink, J. Fritsch, and G. Sagerer*
*Bielefeld University, Faculty of Technology, 33594 Bielefeld, Germany
Email: bwrede@TechFak.Uni-Bielefeld.DE

Abstract— **Current research in robotics is driven by the goal to achieve a high user acceptance of service robots for private households. This implies that robots have to increase their social aptness in order to become a robot companion that is able to interact in a natural way and to carry out tasks for the user.**

In this paper we present the Bielefeld Robot Companion (BIRON) as an example for the development of robot companions. BIRON is a mobile robot that is equipped with an attention system based on a multi-modal sensor system. It can carry out natural speech based dialogs and performs basic functions such as following and learning objects. We argue that the development of robot companions has to be tightly coupled with evaluation phases. We present user studies with BIRON which indicate that the functionality of a robot does not receive as much attention as the natural language interface. This indicates that the communicative behavior of a robot companion is a critical component of the system and needs to be improved before the actual functionalities of the robot can be evaluated and re-designed.

I. Introduction

One of the main current issues in developing interactive autonomous robots is the design of social robots. This focus is motivated by the insight that robots have to exhibit a basic social behavior apart from their functional capabilities in order to be accepted in the environment of a private household. Dauthenhahn and Billard offer a definition of the term *social robots* with respect to the capabilities they exhibit in the interaction with their social environment [5]:

social robots are embodied agents that are part of a heterogeneous group: a society of robots or humans. They are able to recognize each other and engage in social interactions, they possess histories (perceive and interpret the world in terms of their own experience), and they explicitly communicate with and learn from each other.

In order to achieve these goals it is proposed in [6] that a robot has to be able to show the following features and capabilities: Embodiment, emotion, dialog, personality, human-oriented perception, user model, social learning, and intentionality.

Current robotic systems' capabilities are far from showing a human-like level in all these dimensions. However, different aspects have been realized with different degrees of complexity mainly with respect to the features embodiment, human-oriented perception, and dialog.

When comparing the different service robots with respect to these features it becomes apparent that most of them share a similar level of embodiment: the systems are generally based on mobile platforms (e.g. Care-O-bot II [11], CERO [14], HERMES [2], Jijo-2 [1], Lino [16], ROBITA [20]) but only very few have actuators like arms and hands (e.g. Car-O-bot, HERMES) that enable them to fetch and carry objects, which would be one of the fundamental functionalities for a service robot at home. Sensors on such systems generally encompass visual and acoustic (speech) modalities (e.g. Care-O-bot II, HERMES, Jijo-2, Lino, ROBITA, SIG [21]). Thus, despite great differences in their physical appearance current service robots exhibit a rather standardized level of embodiment.

As for human-oriented perception, most systems are able to demonstrate attention-like behavior by visually tracking persons and focusing on a speaking person. Some systems are also able to identify different persons. It is generally observed that this is a crucial basic behavior for robots to gain and keep a person's attention and motivation for interaction.

Less homogenous – and more difficult to compare – are the dialog competences of such robots. It is generally agreed upon that a natural language interface is necessary for easy and intuitive instruction of the robot. However, current dialog systems are often restricted to prototypical command sentences and simple underlying finite state automata. Other modalities than speech, e.g. gestures, are generally ignored.

Emotional perception and production, the development of a personality, building a model of the communication partner, as well as social learning and exhibiting intentionality are features that have partly been demonstrated in so called *sociable robots* (e.g. Kismet [3] or Leonardo [4]) but not on fully autonomous robots that are supposed to fulfill service tasks. However, even such sociable robots do generally not possess sophisticated verbal communication capabilities.

In order to move towards the ambitious goal of a robot companion, which should exhibit both social aptness and service functionalities, it is necessary to perform the development in a closely coupled design-evaluation cycle. In effect, long term user studies such as, for example, performed with CERO are necessary in order to understand the long term influence of contextual variables such as ergonomic features or the

[1]This work has been supported by the European Union within the 'Cognitive Robot Companion' (COGNIRON) project (FP6-IST-002020) and by the German Research Foundation within the Collaborative Research Center 'Situated Artificial Communicators' as well as the Graduate Programs 'Task Oriented Communication' and 'Strategies and Optimization of Behavior'.

reactions of bypassing people. With our robot BIRON we want to address this intersection of social capabilities and functional behavior by enabling the system to carry out a more sophisticated dialog for handling instructions and learning new parts of its environment. One scenario that we envision within the COGNIRON project[1] is a home-tour where a user is supposed to show BIRON around his or her home. This scenario requires BIRON to carry out a natural dialog in order to understand commands e.g. for following and to learn new objects and rooms.

We addressed the issue of evaluation by performing first preliminary user studies in order to evaluate single system components and to better understand in which direction we have to guide the further development of our robot. As we will show, a robot has to reach a certain level of verbal competence before it will be accepted as a social communication partner and before its functional capabilities will be perceived as interesting and useful.

In this paper we will first present the overall system architecture (Section II) and hardware (Section III) before describing the modules in more detail in Sections IV to VI. The current interaction capabilities are shortly described in Section VII. We present results from our user studies Section VIII.

II. System Overview and Architecture

Since interaction with the user is the basic functionality of a robot companion, the integration of interaction components into the architecture is a crucial factor. We propose to use a special control component, the so-called *execution supervisor*, which is located centrally in the robot's architecture [15]. The data flow between all modules is event-based and every message is coded in XML. The modules interact through a specialized communication framework [25]. The robot control system (see Fig. 1) is based on a three-layer architecture [9] which consists of three components: a reactive feedback control mechanism, a reactive plan execution mechanism, and a mechanism for performing deliberative computations.

The execution supervisor, the most important architecture component, represents the reactive plan execution mechanism. It controls the operations of the modules responsible for deliberative computations rather than vice versa. This is contrary to most hybrid architectures where a deliberator continuously generates plans and the reactive plan execution mechanism just has to assure that a plan is executed until a new plan is received. To continuously control the overall system the execution supervisor performs only computations that take a short time relative to the rate of environmental change perceived by the reactive control mechanism.

While the execution supervisor is located in the intermediate layer of the architecture, the dialog manager is part of the deliberative layer. It is responsible for carrying out dialogs to receive instructions given by a human interaction partner. The

[1]COGNIRON is an integrated Project of a European consortium that is supported by the European Union. For more details of this project see http://www.cogniron.org.

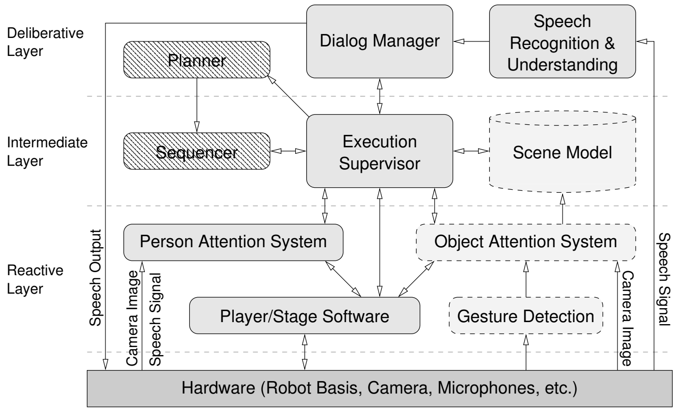

Fig. 1. Overview of the BIRON architecture (implemented modules are drawn with solid lines, modules under development with dashed lines).

dialog manager is capable of managing interaction problems and resolving ambiguities by consulting the user (see Section VI). It receives input from speech processing which is also located on the topmost layer (see Section V) and sends valid instructions to the execution supervisor.

The person attention system represents the reactive feedback control mechanism and is therefore located on the reactive layer (see Section IV). However, the person attention system does not directly control the robot's hardware. This is done by the *Player/Stage* software [10]. *Player* provides a clean and simple interface to the robot's sensors and actuators. Even though we currently use this software to control the hardware directly, the controller can easily be replaced by a more complex component which may be based on, e.g., behaviors.

In addition to the person attention system we are currently developing an object attention system for the reactive layer. The execution supervisor can shift control of the robot from the person attention system to the object attention system in order to focus objects referred to by the user. The object attention will be supported by a gesture detection module which recognizes deictic gestures [13]. Combining spoken instructions and a deictic gesture allows the object attention system to control the robot and the camera in order to acquire visual information of a referenced object. This information will be sent to the scene model in the intermediate layer.

The scene model will store information about objects introduced to the robot for later interactions. This information includes attributes like position, size, and visual information of objects provided by the object attention module. Additional information given by the user is stored in the scene model as well, e.g., a phrase like *"This is my coffee cup"* indicates owner and use of a learned object.

The deliberative layer can be complemented by a component which integrates planning capabilites. This planner is responsible for generating plans for navigation tasks, but can be extended to provide additional planning capabilities which could be necessary for autonomous actions without the human. As the execution supervisor can only handle single commands, a sequencer on the intermediate layer is responsible for decomposing plans provided by the planner. However, in this paper we will focus on the interaction capabilities of the robot.

III. HARDWARE

Our system architecture is implemented on our mobile robot BIRON (see Fig. 2). Its hardware platform is a Pioneer PeopleBot from ActivMedia with an on-board PC (Pentium III, 850 MHz) for controlling the motors and the on-board sensors and for sound processing. An additional PC (Pentium III, 500 MHz) inside the robot is used for image processing and for data association.

The two PCs running Linux are linked by an 100 Mbit Ethernet LAN and the controller PC is equipped with wireless LAN to enable remote control of the robot. As additional interactive device a 12" touch screen display is provided on the front side.

A pan-tilt color camera (Sony EVI-D31) is mounted on top of the robot at a height of 141 cm for acquiring images of the upper body part of humans interacting with the robot. Two AKG far-field microphones which are usually used for hands free telephony are located at the front of the upper platform at a height of 106 cm, right below the touch screen display. The distance between the microphones is 28.1 cm. A SICK laser range finder is mounted at the front at a height of approximately 30 cm.

Fig. 2. BIRON.

IV. THE PERSON ATTENTION SYSTEM

A robot companion should enable users to engage in an interaction as easily as possible. For this reason the robot has to continuously keep track of all persons in its vicinity and must be able to recognize when a person starts talking to it. Therefore, both acoustic and visual data provided by the on-board sensors have to be taken into account: At first the robot needs to know which person is speaking, then it has to recognize whether the speaker is addressing the robot, i.e., looking at it. On BIRON the necessary data is acquired from a multi-modal person tracking framework which is based on *multi-modal anchoring* [8].

A. Multi-Modal Person Tracking

Multi-modal anchoring allows to simultaneously track multiple persons. The framework efficiently integrates data coming from different types of sensors and copes with different spatio-temporal properties of the individual modalities. Person tracking on BIRON is realized using three types of sensors. First, the laser range finder is used to detect humans' legs. Pairs of legs result in a characteristic pattern in range readings and can be easily detected [8]. Second, the camera is used to recognize faces and torsos. Currently, the face detection works for faces in frontal view only [17]. The clothing of the upper body part of a person is observed by tracking the color of the person's torso [7]. Third, the stereo microphones are applied to locate sound sources in front of the robot. By incorporating information from the other cues robust speaker localization is possible [17]. Altogether, the combination of depth, visual, and auditory cues allows the robot to robustly track persons in its vicinity.

However, since BIRON has only limited sensing capabilities – just like a human has only limited cognitive resources – we implemented an attention mechanism for more complex situations with many people moving around BIRON.

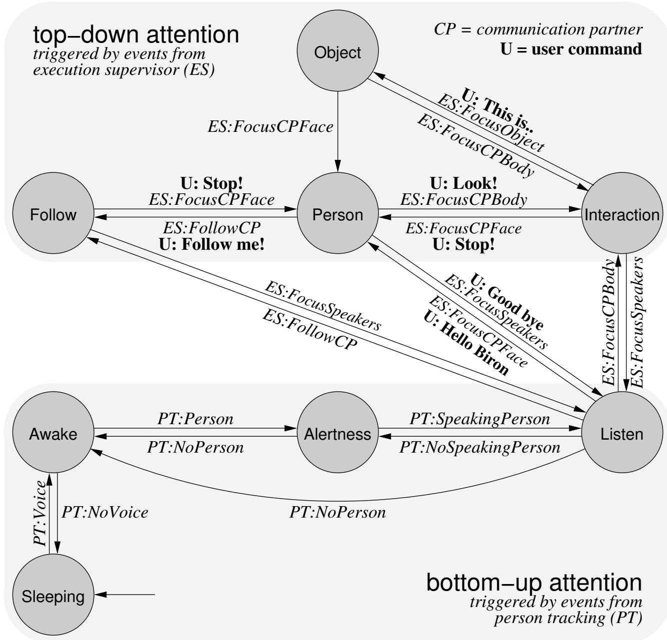

Fig. 3. Finite state machine realizing the different behaviors of the person attention mechanism. Commands from the user, that are processed by the dialog component, are displayed in bold face.

B. Attention Mechanism

The attention mechanism has to fulfill two tasks: On the one hand it has to select the person of interest from the set of observed persons. On the other hand it has to control the alignment of the sensors in order to obtain relevant information from the persons in the robot's vicinity.

The attention mechanism is realized by a finite state machine (see Fig. 3). It consists of several states of attention, which differ in the way the robot behaves, i.e., how the pan-tilt unit of the camera or the robot itself is controlled. The states can be divided into two groups representing *bottom-up attention* while searching for a communication partner and *top-down attention* during interaction.

When bottom-up attention is active, no particular person is selected as the robot's communication partner. The selection of the person of interest as well as transitions between different states of attention solely depend on information provided by the person tracking component. For selecting a person of interest, the observed persons are divided into three categories with increasing degree of relevance. The first category consists of persons that are not speaking. The second category comprises all persons that are speaking, but at the same time are either not looking at the robot or the corresponding decision is not possible, since the person is not in the field of view of the camera. Persons assigned to the third category are of most interest to the robot. These persons are speaking and at the same time are looking at the robot. In this case the robot assumes to be addressed and considers the corresponding person to be a potential communication partner.

Top-down attention is activated as soon as the robot starts to interact with a particular person. During interaction the robot's focus of attention remains on this person even if it is not speaking. Here, in contrast to bottom-up attention, transitions between different states of attention are solely triggered by the execution supervisor which reacts to user commands processed by the dialog component. For detailed information concerning the control of the hardware see [12].

V. Speech Processing

As speech is the most important modality for a multi-modal dialog, speech processing has to be done thoroughly. On BIRON there are two major challenges: Speech recognition has to be performed on distant speech data recorded by the two on-board microphones and speech understanding has to deal with spontaneous speech.

While the recognition of distant speech with our two microphones is achieved by beam-forming [18], the activation of speech recognition is controlled by the attention mechanism presented in the previous section. Only if a tracked person is speaking and looking at the robot at the same time, speech recognition and understanding takes place. Since the position of the speaker relative to the robot is known from the person tracking component, the time delay can be estimated and taken into account for the beam-forming process.

The speech understanding component processes recognized speech and has to deal with spontaneous speech phenomena. For example, large pauses and incomplete utterances can occur in such task oriented and embodied communication. However, missing information in an utterance can often be acquired from the scene. For example the utterance *"Look at this"* and a pointing gesture to the table can be combined to form the meaning *"Look at the table"*. Moreover, fast extraction of semantic information is important for achieving adequate response times.

We obtain fast and robust speech processing by combining the speech understanding component with the speech recognition system. For this purpose, we integrate a robust LR(1)-parser into the speech recognizer as proposed in [24]. Besides, we use a semantic-based grammar which is used to extract instructions and corresponding information from the speech input. A semantic interpreter forms the results of the parser into frame-based XML-structures and transfers them to the dialog manager. Hints in the utterances about gestures are also incorporated. For our purpose, we consider co-verbal gestures only.

For the object attention system it is intended to use this information in order to detect a specified object. Thus, this approach supports the object attention system and helps to resolve potential ambiguities.

VI. Dialog

The model of the dialog manager is based on a set of *finite state machines* (FSM), where each FSM represents a specific dialog [23]. The FSMs are extended with the ability of recursive activation of other FSMs and the execution of an action in each state. Actions that can be taken in certain states are specified in the *policy* of the dialog manager. These actions include the generation of speech output and sending events like orders and requests to the execution supervisor. The dialog *strategy* is based on the so-called *slot-filling* method [22]. The task of the dialog manager is to fill enough slots to meet the current dialog goal, which is defined as a goal state in the corresponding FSM. The slots are filled with information coming from the user and other components of the robot system. After executing an action, which is determined by a lookup in the dialog policy, the dialog manager waits for new input from the execution supervisor or the speech understanding system.

As users interacting with a robot companion often switch between different contexts, the slot-filling technique alone is not sufficient for adequate dialog management. Therefore, the processing of a certain dialog can be interrupted by another one, which makes alternating instruction processing possible. Dialogs are specified using a declarative definition language and encoded in XML in a modular way. This increases the portability of the dialog manager and allows an easier configuration and extension of the defined dialogs.

VII. Interaction Capabilities

In the following we describe the interaction capabilities BIRON offers to the user in our current implementation. Initially, the robot observes its environment. If persons are present in the robot's vicinity, it focuses on the most interesting one. A user can start an interaction by greeting the robot with, e.g., *"Hello BIRON"* (see Fig. 3). Then, the robot keeps this user in its focus and can not be distracted by other persons talking. Next, the user can ask the robot to follow him to another place in order to introduce it to new objects. While the robot follows a person it tries to maintain a constant distance to the user and informs the person if she moves too fast. When the robot reaches a desired position the user can instruct it to stop. Then, the user can ask the robot to learn new objects. In this case the camera is lowered to also get the hands of the user in the field of view. When the user points to a position and gives spoken information like *"This is my favorite cup"*, the object attention system is activated in order to center the referred object. However, since the gesture recognition and the object attention modules are not yet integrated in our system, this behavior is simulated by always moving the camera to a

Fig. 4. Several scenes from users interacting with BIRON during our first user studies.

predefined position when reaching the attentional state *Object*. If the user says *"Good-bye"* to the robot or simply leaves while the robot is not following the user, the robot assumes that the current interaction is completed and looks around for new potential communication partners.

VIII. Evaluation

We carried out first user studies with BIRON by assessing qualitative statements from users about the capabilities of BIRON. We asked 21 subjects to interact with BIRON. Figure 4 shows some interaction scenes from these experiments. Interaction times (i.e. the time where only one user interacted with BIRON) averaged between 3 and 5 minutes. As an introduction the users were given an overview of BIRON's interaction capabilities which displayed a schema of potential commands similar to the graph shown in Figure 3. Afterwards they had to fill out a questionnaire where we asked, among others, for the most and the least preferred features that they had experienced during their interactions with BIRON. More detailed results of this evaluation are reported [19].

It turned out that the most interesting features for users were the natural language interface and the person attention behavior (see Fig. 5). The more task-oriented functions – the following behavior and the object learning ability – received less positive feedback. This indicates that the functional capabilities of BIRON did not receive as much attention as one would expect and seem to be obscured by other features of the system.

On the other hand, although all users did already have some experience with speech recognition systems (ASR), the most frequently named dissatisfaction concerned the errors of the ASR system (see Fig. 6). Wishes for a more flexible dialog and a more stable system were the only other significant dimensions of answers to this open question, although less frequently named.

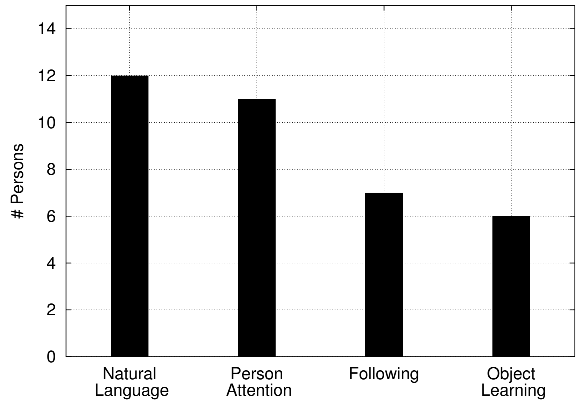

Fig. 5. User answers to the question "What did you like most?"

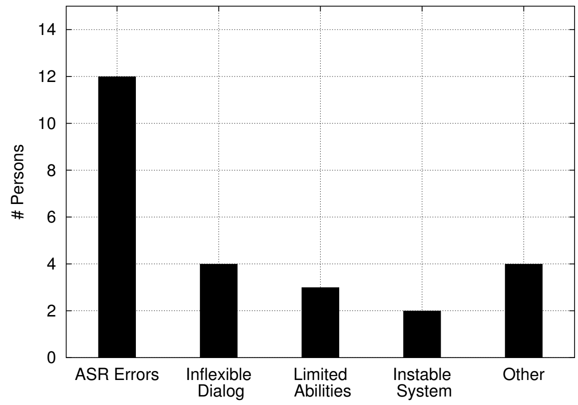

Fig. 6. User answers to the question "What did you like least?"

These results emphasize the importance of a natural language interface which allows for natural interactions. However, they also demonstrate that users are extremely sensitive to problems that occur within the communication. Thus, the natural language capability of a robot is a crucial part for human-robot interaction. If the communication does not proceed in a smooth way, the user will not be motivated to access all the potential functionalities of the robot.

In addition to these results we also assessed the usefulness of the feedback of different internal processing results and states. It turned out that users generally found feedback very helpful. However, users tend to have highly individual preferences as to the means of feedback they prefer. While some users liked to see the results of the ASR system, others found these too technical and disturbing from the actual task. On the other hand, the feedback of the internal attentional state of the system was generally perceived as very helpful. This shows that while feedback on the internal system status is helpful it has to be conveyed in an acceptable way to the user. A powerful means that humans use in their communication are nonverbal signals such as gestures or mimic. It seems to be promising to implement more of such nonverbal communication on a robotic companion as demonstrated on sociable robots such as Kismet or Leonardo ([3], [4]).

IX. Conclusion

In order for a robot to be accepted as a social communication partner it should exhibit a range of features and functionalities. The main features that current state-of-the-art robots exhibit concern embodiment, human-oriented perception and dialog.

In this paper we argued that the levels of embodiment and human-oriented perception, that current state-of-the-art robots share, have reached a standard which is – with the exception of missing actuators – quite acceptable for human users. We demonstrated this with first user studies on BIRON which showed that the attentional behavior of BIRON receives significant positive feedback while the functional features (person following, object learning) did not receive as much attention by the same subjects. We suppose that this is due to the limitations of the natural language interface which, while being the preferred communication channel for human users, is currently the most critical system component. Here, user wishes direct our research towards a more robust speech recognition system and a more flexible dialog. We are currently planning to use a head-mounted microphone for getting cleaner speech for the speech recognition system in addition to the stereo microphones that we use for the speaker localization.

These results indicate that a robot companion has to show acceptable communication skills in order to be acceptable both at a social and a functional level. They also demonstrate that it is necessary to tightly couple user studies with design and development phases. In order to build robots that are acceptable as social communication partners it is necessary to identify critical aspects of the system. Within the design-development-evaluation cycle of BIRON the current findings direct our research towards developing new means for a more robust, embodied communication framework.

References

[1] H. Asoh, Y. Motomura, F. Asano, I. Hara, S. Hayamizu, K. Itou, T. Kurita, T. Matsui, N. Vlassis, R. Bunschoten, and B. Kröse. Jijo-2: An office robot that communicates and learns. *IEEE Intelligent Systems*, 16(5):46–55, 2001.

[2] R. Bischoff and V. Graefe. Demonstrating the humanoid robot *HERMES* at an exhibition: A long-term dependability test. In *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems; Workshop on Robots at Exhibitions*, Lausanne, Switzerland, 2002.

[3] C. Breazeal. *Designing Sociable Robots*. Bradford Books, 2002.

[4] C. Breazeal, D. Buchsbaum, J. Gray, D. Gatenby, and B. Blumberg. Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots. *Artificial Life*, 2004. to appear.

[5] K. Dautenhahn and A. Billard. Bringing up robots or – the psychology of socially intelligent robots: From theory to implementation. In *Proc. of the Autonomous Agents*, 1999.

[6] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. *Robotics and Autonomous systems*, 42:143–166, 2003.

[7] J. Fritsch, M. Kleinehagenbrock, S. Lang, G. A. Fink, and G. Sagerer. Audiovisual person tracking with a mobile robot. In *Proc. Int. Conf. on Intelligent Autonomous Systems*, pages 898–906. IOS Press, 2004.

[8] J. Fritsch, M. Kleinehagenbrock, S. Lang, T. Plötz, G. A. Fink, and G. Sagerer. Multi-modal anchoring for human-robot-interaction. *Robotics and Autonomous Systems, Special issue on Anchoring Symbols to Sensor Data in Single and Multiple Robot Systems*, 43(2–3):133–147, 2003.

[9] E. Gat. On three-layer architectures. In D. Kortenkamp, R. P. Bonasso, and R. Murphy, editors, *Artificial Intelligence and Mobile Robots: Case Studies of Successful Robot Systems*, chapter 8, pages 195–210. MIT Press, Cambridge, MA, 1998.

[10] B. P. Gerkey, R. T. Vaughan, and A. Howard. The player/stage project: Tools for multi-robot and distributed sensor systems. In *Proc. Int. Conf. on Advanced Robotics*, pages 317–323, 2003.

[11] B. Graf, M. Hans, and R. D. Schraft. Care-O-bot II—Development of a next generation robotic home assistant. *Autonomous Robots*, 16(2):193–205, 2004.

[12] A. Haasch, S. Hohenner, S. Hüwel, M. Kleinehagenbrock, S. Lang, I. Toptsis, G. A. Fink, J. Fritsch, B. Wrede, and G. Sagerer. BIRON – The Bielefeld Robot Companion. In *Proc. Int. Workshop on Advances in Service Robotics*, pages 27–32, 2004.

[13] N. Hofemann, J. Fritsch, and G. Sagerer. Recognition of deictic gestures with context. In *Proc. DAGM'04*. Springer-Verlag, 2004. to appear.

[14] H. Hüttenrauch and K. Severinson Eklundh. Fetch-and-carry with CERO: Observations from a long-term user study with a service robot. In *Proc. IEEE Int. Workshop on Robot-Human Interactive Communication (ROMAN)*, pages 158–163. IEEE Press, 2002.

[15] M. Kleinehagenbrock, J. Fritsch, and G. Sagerer. Supporting advanced interaction capabilities on a mobile robot with a flexible control system. In *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Sendai, Japan, September/October 2004. to appear.

[16] B. J. A. Kröse, J. M. Porta, A. J. N. van Breemen, K. Crucq, M. Nuttin, and E. Demeester. Lino, the user-interface robot. In *European Symposium on Ambient Intelligence (EUSAI)*, pages 264–274, 2003.

[17] S. Lang, M. Kleinehagenbrock, S. Hohenner, J. Fritsch, G. A. Fink, and G. Sagerer. Providing the basis for human-robot-interaction: A multi-modal attention system for a mobile robot. In *Proc. Int. Conf. on Multimodal Interfaces*, pages 28–35. ACM, 2003.

[18] S. J. Leese. Microphone arrays. In G. M. Davis, editor, *Noise Reduction in Speech Applications*, pages 179–197. CRC Press, Boca Raton, London, New York, Washington D.C., 2002.

[19] S. Li, M. Kleinehagenbrock, J. Fritsch, B. Wrede, and G. Sagerer. "BIRON, let me show you something": Evaluating the interaction with a robot companion. In *Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics, Special Session on Human-Robot Interaction*, The Hague, The Netherlands, October 2004. IEEE. to appear.

[20] Y. Matsusaka, T. Tojo, and T. Kobayashi. Conversation robot participating in group conversation. *IEICE Trans. on Information and System*, E86-D(1):26–36, 2003.

[21] H. G. Okuno, K. Nakadai, and H. Kitano. Social interaction of humanoid robot based on audio-visual tracking. In *Proc. Int. Conf. on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, Cairns, Australia, 2002. Lecture Notes in Artificial Intelligence, Springer.

[22] B. Souvignier, A.Kellner, B. Rueber, H. Schramm, and F. Seide. The thoughtful elephant: Strategies for spoken dialog systems. In *IEEE Trans. on Speech and Audio Processing*, volume 8, pages 51–62, 2000.

[23] I. Toptsis, S. Li, B. Wrede, and G. A. Fink. A multi-modal dialog system for a mobile robot. In *Proc. Int. Conf. on Spoken Language Processing*, 2004. to appear.

[24] S. Wachsmuth, G. A. Fink, and G. Sagerer. Integration of parsing and incremental speech recognition. In *Proc. European Conf. on Signal Processing*, volume 1, pages 371–375, Rhodes, 1998.

[25] S. Wrede, J. Fritsch, C. Bauckhage, and G. Sagerer. An XML based framework for cognitive vision architectures. In *Proc. Int. Conf. on Pattern Recognition*, Cambridge, UK, 2004. to appear.