ToBI - Team of Bielefeld: The Human-Robot Interaction System for RoboCup@Home 2009

Sven Wachsmuth, Marc Hanheide, Frederic Siepmann, Thorsten Spexard

Faculty of Technology, Bielefeld University,
Universitätstraße 25, 33615 Bielefeld, Germany

Abstract. The ToBI robocup team has been newly founded in Jan 2009 in order to proceed existing long-term research in the development of robot companions for domestic environments towards new challenges in more standardized benchmarking procedures, like RoboCup@Home. The main features of the ToBI system are a flexibile *Active Memory-based* architecture that enables the fast integration of new processing modules and new system behaviors and the modeling of mixed-initiative strategies for multi-modal dialog. The overall goal is an *out-of-the-box* robot that is able to successfully interact with naïve users. In this paper we describe the technical basis on which the ToBI system is based and give some insights on previous evaluation experiences.

1 Introduction

The RoboCup@Home competition aims at bringing robotic platforms into people's every-day lives in their own home environments. This includes many challenges from the autonomous navigation and acting in regular apartments, the recognition and learning of specific places, objects, and situations, as well as the social interaction with people in the apartment. The last aspect is inevitable and particularly challenging if the robot is interacting with naïve users. Studies in human-robot interaction have shown that the successful completion of cooperative tasks, like showing the robot around in the own apartment or teaching the robot new objects, critically depend on appropriate feedback cues [1, 2]. The robot also needs to decide when to take the own initiative in a dialog if the user is confused. The more general perspective on such kind of feedback loops is the understanding of human-robot interaction as a *social task*. In order to enable a robot to deal with flexible real-world environments it needs to overcome pre-programmed behaviors and fixed scenarios. However, classical paradigms from machine learning do not apply to the scenario of a user who is continuously interacting with the robot. It requires novel paradigms that go beyond defined training and test sets, but demand for a continuous validation and adaptation of acquired models, embedded in the interaction itself. This embedding of learning has recently been mentioned as “socially guided machine learning” [3]. As a consequence, training and testing are not seperated anymore and the robot needs to provide out-of-the-box skills that enable a learning of a specific environment by interacting.

The ToBI system that is specifically adapted for the RoboCup@Home scenarios is based on technology of the Bielefeld Robot Companion (BIRON) that has been developed over the last 7 years [4–6]. In the following sections, we will describe the hardware and software basis of the robot, briefly explain different skills, and discuss some of the learned lessons in previous evaluation studies.

2 Hardware

Fig. 1. The robot ToBI with it's components shown on the right. From top right: Pan-/Tilt-camera, interfacial microphone, Pioneer 5DOF arm and laser range finder.

The robot platform *ToBI* is based on the research platform *GuiaBot*™ by MobileRobots[1] customized and equipped with sensors that allow analysis of the current situation in a human-robot interaction (HRI). ToBI is a consequent advancement of the *BIRON* (**BI**lefeld **R**obot compan**ION**) platform, which has been under continuous development since seven years. It comprises two piggyback laptops to provide the computational power and to achieve a system running autonomously and in real-time for HRI.

The robot base is a PatrolBot™which is 59cm in length, 48cm in width, and 38cm in height, weighs approx. 45 kilograms with batteries and is maneuverable with 1.7 meters per second maximum translation and 300+ degrees rotation per second. The drive is a two-wheel differential drive with two passive rear casters for balance. Its solid foam-filled 19cm diameter wheels are at the center of rotation. It can climb up to a 20 percent grade, over a 2cm sill and carry a 12 kilogram payload. Inside the base there is a 180 degree laser range finder (SICK LMS200, see Fig.1 bottom right). It can sense objects as far away as 50 meters with a ranging accuracy of 18 millimeters at a distance of up to 18m. The scanning height is at 30cm above the floor. In contrast to most other PatrolBot bases, ToBI does not use an additional internal computer. The piggyback laptops are equipped with Intel Core2Duo©2GB main memory and are running Linux. The camera used here is a 12x zoom pan-/tilt camera (SONY PTZ, see Fig.1 top right) that is able to scan an area of ±100 degree in front of the robot. For localization of sound direction two interfacial microphones are mounted on the top of the robot's body (see Fig.1 second from top on the right). For the detection of planar surfaces and obstacles that are not in the laser range ToBI is equipped with an

[1] www.mobilerobots.com

optical imaging system for real time 3D image data acquisition (SwissRanger). The camera features an integrated, modulated infrared light source that enables a time-of-flight based measurement of a real time depth map.

Additionally the robot is equipped with a Pioneer 5 degrees-of-freedom (DOF) arm (see Fig.1 second from bottom on the right); a small and lightweight manipulator driven by six open-loop servo motors. The Pioneer arms end-effector is a gripper whose foam-lined fingers allowing to grasp and manipulate objects as large as a can and as heavy as 150 grams throughout the arms envelope of operation.

The upper part of the robot's body houses a touch screen ($\approx 15in$) as well as the system speaker. The overall height of the robot ToBI is approximately 130cm.

3 System Architecture

For complex tasks as targeted by the RoboCup@Home challenge many different software components are in use that have to be orchestrated, coordinated, and integrated into one system. In order to provide the required flexibility, a cognitively motivated memory architecture serves as the foundation of our robotic system – both for the functional system architecture as also for the software architecture.

3.1 The Active Memory for Information-driven Integration

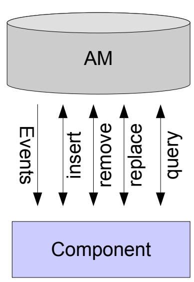

Fig. 2.

We proposed the concept of an *active memory* (AM) for building cognitive systems already some years ago. It basically puts forward the concept of event-driven integration (EDI) on the basis of flexible event notification and XML-based representations as a document-oriented data model. In particular it comprises an "active memory service" (AMS) as a central integration broker for coordination and shared data management. Conceptually, all information generated and revised by components in the system is mediated through this active memory, where it can persistently be stored and retrieved from. In order to tackle performance issues and allow flexible, content-based event subscriptions, an advanced *Filtering, Transformation, and Selection (FTS)* concept [7] is applied to facilitate high responsitivity on all levels of the robot's architecture. Components can subscribe for particular fragments of information by means of XPath subscriptions and get notified whenever this subscribed memory content changes or shows up in the memory. Hence, operations that modify the memory's content are *insert*, *remove*, *replace* as also illustrated in Fig. 2. Together with an additional *query* operation, these resemble functionalities of database systems. Components in this concept are termed *memory processes* that generate, receive, and interpret information and are locally triggered by the exchange of the information atoms,

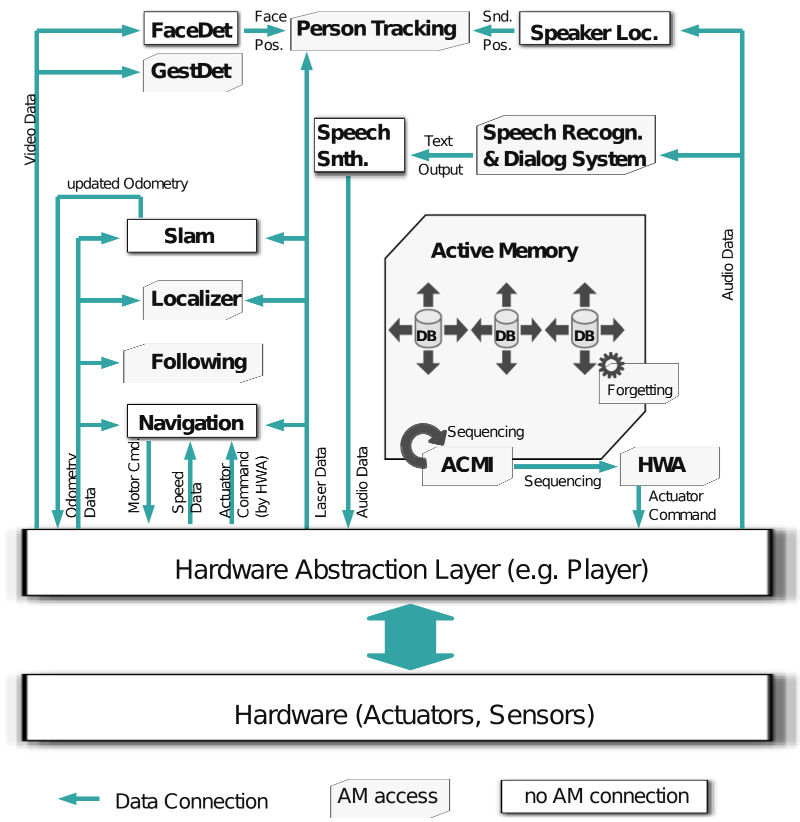

Fig. 3. A sketch of the general system architecture of the ToBI robot. The Active Memory serves as a global data repository and coordination mechanism where intermediate processing results are stored and various events are triggered. The Active Control Memory Interface (ACMI) sequences the robot actuation based on the memory content. The Hardware Arbitration (HWA) provides the interface to the Hardware Abstraction Layer.

so-called *memory elements* through respective subscriptions. The AMS itself can be seen as a global tupel-space, providing coordinated read-write access to all memory processes, but going beyond traditional architectures, e.g., by means of the above mentioned FTS concepts. For more details regarding the principles of active memories and their application in more general intelligent systems refer to [8]. This AM concept constitutes the foundation of the architecture of the ToBI robot. Its practical use has been already shown with the robot BIRON in a home-tour scenario, where the user introduces an apartment unknown before to the robot. The experiments conducted in this scenario especially focused on flexible coordination and dialog strategies applying production-rule systems (ACMI) for coordination and a grouding-based dialog system [9, 10]. The shared architectural concepts of the ToBI and BIRON robots are illustrated in Fig. 3.

From an engineering point of view, the event-driven AM concept is directly supported by the Open-Source integration framework *XCF* [2][11]. Besides an efficient implementation of the event-driven integration concepts it furthermore supports common interaction patterns such as request-reply and publish-subscribe across computational hosts to allow distributed, parallel, and asynchronous processing. It features adapters for the open-source computer vision framework *icewing* [3] employed for image processing in ToBI, and common robotic toolkits such as Player/Stage, MRPT, and others more.

3.2 An Abstraction API for Advanced Behavior-oriented Design

While the AM is our basic architectural priniciple that is applied for the coordination and integration, often an abstraction layer is required in order to design appropriate behaviors on a task and scenario level. In particular for evaluation, HRI studies, training and teaching purposes the *BonSAI (BirON Sensor Actuator Interface)*-API has been developed. The goal is to provide a very robust and easy to use abstraction layer comprising a comprehensive set of advanced abilities of the robot accessible through a well-designed Java API[4] for simple implementation of robot behaviors, picking up concepts of behavior-oriented design [12]. Its basic conceptual abstraction is to hide the complexity of advanced sensor compounds and actuator controllers behind abstract sensors and actuators with a minimalistic API. The backend of the API built upon the AM concepts takes care of location transparency, handling of problems resulting from distributed computation, and itself encapsulates control loops, e.g. for motion control. Hence, the BonSAI-API features abstract high-level sensors such as a `PersonSensor`, which detects and tracks all persons in the vicinity of the robot fusing different perceptual cues, and actuators such as a `NavigationActuator` comprising motion control, obstacle avoidance, and path planning accessible to a simple interface. The BonSAI-API allows to link these advanced sensors and actuators into closed-loop behaviors on a high level of abstraction, also according to specific tasks of the RoboCup@Home challenge.

4 Software Components

Although the software architecture is organized in terms of components or *memory processes*, the functional skills of the system emerge from the interaction of multiple components. Therefore, the following description is organized into skills rather than single components. These constitute also respective sensors and actuators in the BonSAI-API. Though many more are present in the system, three main skill are detailed in the following: Person Tracking, Dialog System, and Localization with Navigation.

[2] http://xcf.sf.net
[3] http://icewing.sf.net
[4] see https://code.ai.techfak.uni-bielefeld.de/bonsai for documentation of the current BonSAI API

4.1 Person Tracking

As a basis for any kind of human-robot-interaction a robust detection and tracking of interaction partners is essential. The applied person tracking approach [6] uses multiple cues from different modalities. First, data of the laser range finder is used to identify pairs of legs in a 180 degree area in front of the robot. Additionally the pan-tilt camera is used in combination with a face detector. Since the camera field of view is too small to observe the whole area in front of the robot, it is actively controlled by considering hypotheses generated from laser data to search for people's faces. Face recognition using Active Appearance Models and SVM is available for person recognition [13]. Speaker localization is supported by a pair of stereo microphones applying signal delay computation. These three cues are used to determine a person position and track her according the *anchoring* principle . In order to handle multiple interaction partners, an attention control decides which person might be the current interaction partner based on movement, gaze, and talkativeness. Pointing gestures can guide the robots attention to certain areas in the environment [14].

4.2 Dialog System

After a communication partner has been successfully detected, the interaction is guided by a powerful naturally spoken dialog system [15]. The dialog system is based on the grounding of multi-modal interaction units. It gets information about the current system state from the memory and combines the information with the user utterances processed by a speaker-independent speech recognizer and speech understanding system. Combining user input and system state the dialog provides commands to the system as well as feedback to the user via speech synthesis (using Mary open-source software[5]).

4.3 Localization and Navigation

Being able to detect people and to communicate in a natural way by speech and gesture, the demonstrator has to autonomously act in its vicinity. A first step towards this ability is to acquire a model of its vicinity by taking a user guided tour. Therefore the robot uses the *Following* behavior [16]. One can assume that a human guide uses an optimal way, which contains only few obstacles. However, to avoid fixed or dynamic obstacles like other people crossing the way, and for estimating the direction from one route point to another one, an open source navigation component[6] based on the bubble band approach with integrated obstacle avoidance is used. During the guided tour SLAM is used in combination with data from the laser range finder and wheel odometry to create a map. The implemented interactive location learning [10] integrates symbolic information like room or object labels into human-augmented maps to facilitate autonomous navigation afterwards.

[5] http://mary.dfki.de
[6] http://libsunflower.sf.net

	User Trials N1	User Trials N2
subjects	10	14
task completed successfully	9	13
mean duration of task	08:20	10:54
std. dev. duration of task	2.46	4.52

Table 1. Statistics of user trials (duration in min.). More details to be found in [1].

5 Evaluation, Experiences, and Conclusion

The presented components, hardware, and architectural concepts have been applied in numerous studies with different foci using the robot BIRON. Our goal in these studies was to establish an evaluation-development cycle to iteratively improve the interaction capabilities of our robots. The focus of our research here lies in particular on the interaction design and adaptive interaction skills enabling naïve users to successfully interact with the root out-of-the-box. In the core of this endeavour lies *Systemic Interaction Analysis (SInA)* [1], applying task analysis of naïve user trials. Hence, the central abilities and functions of the ToBI robot have been already improved iteratively.

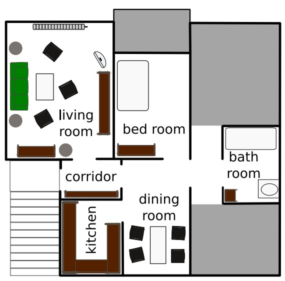

Fig. 4. Apartment.

The investigated tasks in these studies share many commonalities with the RoboCup@Home challenges and were studied in a real off-campus apartment with typical narrow passages and obstructions sketched in Fig. 4. Table 1 summarizes time measurements of two successive trials (N1 & N2) carried out with the robot BIRON for a task rather similar to some of the RoboCup@Home challenges. The instruction to the users, who had never interacted with a robot before, was to (i) establish contact with the robot and attract its attention to accept you as an interaction partner, (ii) guide the robot through the cluttered apartment; from the living room to the dining room via the corridor, and (iii) to show (by guidance and gestures) and label the living room, the green armchair (in the living room), the dining room, and the dining room table. Almost all naïve users managed to conduct the given task in reasonable time. The particular tasks of RoboCup@Home are more specific, so that the mentioned durations are divided for several smaller tasks.

In this paper, we presented the basic hardware and software concepts and components of the ToBI robot to participate in the RoboCup@Home competition. The experiences and achievements gathered with the BIRON platform serve as the foundation for the ToBI robot. Its general ability to successfully conduct tasks as envisioned by the competition has been proven by several studies also with lay persons in real apartments and not only lab settings. The memory-centered architectural concept of ToBI assures a rather general robot system that is not only engineered to conduct the tasks of RoboCup@Home but provide more general abilities of an interactive robot, in particular with respect to learning and interaction.

References

1. Lohse, M., Hanheide, M., Rohlfing, K., Sagerer, G.: Systemic Interaction Analysis (SInA) in HRI. In: Conference on Human-Robot Interaction (HRI), San Diego, CA, USA, IEEE, IEEE (11/03/2009 2009)
2. Lütkebohle, I., Peltason, J., Schillingmann, L., Elbrechter, C., Wrede, B., Wachsmuth, S., Haschke, R.: The curious robot - structuring interactive robot learning. In: International Conference on Robotics and Automation, Kobe, Japan, IEEE, IEEE (14/05/2009 2009)
3. Thomaz, A.L., Hoffman, G., Breazeal, C.: Experiments in socially guided machine learning: understanding how humans teach. In: Proc. Conf on Human-robot interaction, ACM (2006) 359–360
4. Haasch, A., Hohenner, S., Hwel, S., Kleinehagenbrock, M., Lang, S., Toptsis, I., Fink, G.A., Fritsch, J., Wrede, B., Sagerer, G.: Biron – the bielefeld robot companion. In: Proc. Int. Workshop on Advances in Service Robotics. (2004) 27–32
5. Wrede, B., Kleinehagenbrock, M., Fritsch, J.: Towards an integrated robotic system for interactive learning in a social context. In: Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems - IROS 2006, Bejing (2006)
6. Hanheide, M., Sagerer, G.: Active memory-based interaction strategies for learning-enabling behaviors. In: International Symposium on Robot and Human Interactive Communication (RO-MAN), Munich (01/08/2008 2008)
7. Lütkebohle, I., Schäfer, J., Wrede, S.: Facilitating re-use by design: A filtering, transformation, and selection architecture for robotic software systems. In: ICRA Workshop on Software Development in Robotics (SDIR). (2009) submitted.
8. Wachsmuth, S., Wrede, S., Hanheide, M.: Coordinating interactive vision behaviors for cognitive assistance. Computer Vision and Image Understanding **108**(1–2) (October 2007) 135–149
9. Spexard, T.P., Siepmann, F.H.K., Sagerer, G.: A memory-based software integration for development in autonomous robotics. In: Proc. Int. Conf. on Intelligent Autonomous Systems, Baden-Baden, Germany (23/07/2008 2008) 49–53
10. Peltason, J., Siepmann, F.H., Thorsten P. Spexard, B.W., Hanheide, M., Topp, E.A.: Mixed-initiative in human augmented mapping. In: Processings Int. Conference on Robotics and Automation. (2009) to be published.
11. Fritsch, J., Wrede, S.: An integration framework for developing interactive robots. In Brugali, D., ed.: Springer Tracts in Advanced Robotics. Volume 30. Springer, Berlin (2007) 291–305
12. Bryson, J.: The Behavior-Oriented Design of Modular Agent Intelligence. In: Agent Technologies, Infrastructures, Tools, and Applications for E-Services. Volume 2592 of Lecture Notes in Computer Science. Springer (2003) 61–76
13. Hanheide, M., Wrede, S., Lang, C., Sagerer, G.: Who am i talking with? a face memory for social robots. In: Proc. Int. Conf. on Robotics and Automation, Pasadena, CA, USA, IEEE (2008) to appear
14. Haasch, A., Hofemann, N., Fritsch, J., Sagerer, G.: A multi-modal object attention system for a mobile robot. In: Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, Edmonton, Alberta, Canada, IEEE, IEEE (August 2005) 1499–1504
15. Li, S., Wrede, B., Sagerer, G.: A computational model of multi-modal grounding. In: Proc. ACL SIGdial workshop on discourse and dialog, in conjunction with COLING/ACL 2006, ACL Press, ACL Press (2006) 153–160
16. Yuan, F., Hanheide, M., Sagerer, G.: Spatial context-aware person-following for a domestic robot. In: International Workshop on Cognition for Technical Systems, Munich, Germany (2008)