We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Towards Tutoring an Interactive Robot

Britta Wrede, Katharina J. Rohlfing, Thorsten P. Spexard and Jannik Fritsch
Bielefeld University, Applied Computer Science Group
Germany

1. Introduction

Many classical approaches developed so far for learning in a human-robot interaction setting have focussed on rather low level motor learning by imitation. Some doubts, however, have been casted on whether with this approach higher level functioning will be achieved (Gergeley, 2003). Higher level processes include, for example, the cognitive capability to assign meaning to actions in order to learn from the tutor. Such capabilities involve that an agent not only needs to be able to mimic the motoric movement of the action performed by the tutor. Rather, it understands the constraints, the means, and the goal(s) of an action in the course of its learning process. Further support for this hypothesis comes from parent-infant instructions where it has been observed that parents are very sensitive and adaptive tutors who modify their behaviour to the cognitive needs of their infant (Brand et al., 2002).

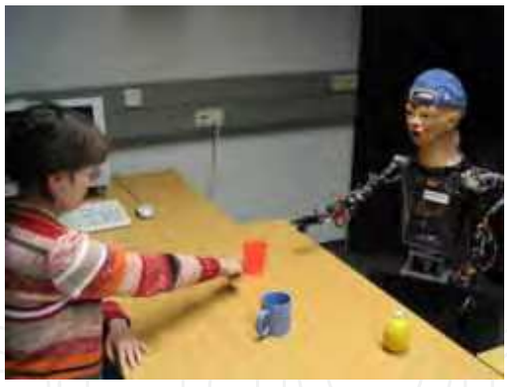


Figure 1. Imitation of deictic gestures for referencing on the same object

Based on these insights, we have started our research agenda on analysing and modelling learning in a communicative situation by analysing parent-infant instruction scenarios with automatic methods (Rohlfing et al., 2006). Results confirm the well known observation that parents modify their behaviour when interacting with their infant. We assume that these modifications do not only serve to keep the infant's attention but do indeed help the infant to understand the actual goal of an action including relevant information such as constraints and means by enabling it to structure the action into smaller, meaningful chunks. We were

Source: Humanoid Robots: Human-like Machines, Book edited by: Matthias Hackel
ISBN 978-3-902613-07-3, pp. 642, Itech, Vienna, Austria, June 2007

able to determine first objective measurements from video as well as audio streams that can serve as cues for this information in order to facilitate learning of actions.
Our research goal is to implement such a mechanism on a robot. Our robot platform *Barthoc* (Bielefeld Anthropomorphic RoboT for Human-Oriented Communication) (Hackel et al., 2006) has a human-like appearance and can engage in human-like interactions. It encompasses a basic attention system that allows it to focus the attention on a human interaction partner, thus maintaining the system's attention on the tutor. Subsequently, it can engage in a grounding-based dialog to facilitate human robot interaction.
Based on our findings on learning in parent-infant interaction and Barthoc's functionality as described in this Chapter, our next step will be to integrate algorithms for detecting infant-directed actions that help the system to decide when to learn and when to stop learning (see Fig. 1). Furthermore, we will use prosodic measures and correlate them with the observed hand movements in order to help structuring the demonstrated action. By implementing our model of communication-based action acquisition on the robot-platform Barthoc we will be able to study the effects of tutoring in detail and to enhance our understanding of the interplay between representation and communication.

2. Related Work

The work plan of social robotics for the next future is to create a robot that can observe a task performed by a human (Kuniyoshi et al., 1994) and interpret the observed motor pattern as a meaningful behaviour in such a manner that the meanings or goals of actions can activate a motor program within the robot.
Within the teaching by showing paradigm (Kuniyoshi et al., 1994), the first step according to this work plan has been accomplished by focussing on mapping motor actions. Research has been done on perception and formation of internal representation of the actions that the robot perceives (Kuniyoshi et al., 1994), (Wermter et al., 2005). However, from the ongoing research we know that one of the greatest challenges for robotics is how to design the competence not only of imitating but of action understanding. From a developmental psychology perspective Gergely (2003) has pointed out that the so far pursued notion of learning lacks higher-level processes that include "understanding" of the semantics in terms of goal, means and constraints. What is meant by this critique is the point that robots learning from human partners not only should know *how* to imitate (Breazeal et al., 2002) (Demiris et al., 1996) and *when* to imitate (Fritsch et al., 2005) but should be able to come up with their own way of reproducing the achieved change of state in the environment. This challenge, however, is tightly linked to another challenge, occurring exactly because of the high degree of freedom of how to achieve a goal. This forms the complexity of human actions, and the robot has to cope with action variations, which means that when comparing across subjects, most actions typically appear variable at a level of task instruction. In other words, we believe that the invariants of action, which are the holy grail of action learning, will not be discovered by analyzing the "appearance" of a demonstrated action but only by looking at the higher level of semantics. One modality that is pre-destined for analyzing semantics is speech. We therefore advocate the use of multiple modalities, including speech, in order to derive the semantics of actions.
So far these points have barely been addressed in robotics: Learning of robots usually consists of uni-modal abstract learning scenarios involving generally the use of vision systems to track movements and transform observed movements to ones own morphology

("imitation"). In order for a robot to be able to learn from actions based on the imitation paradigm, it seems to be necessary to reduce the variability to a minimum, for example by providing another robot as a teacher (Weber et al., 2004).
We argue that information from communication, such as the coordination of speech and movements or actions, in learning situations with a human teacher can lighten the burden of semantics by providing an organization of presented actions.

3. Results from Parent-infant tutoring

In previous work (Rohlfing et al., 2006) we have shown that in parent-child interaction there is indeed a wealth of cues that can help to structure action and to assign meaning to different parts of the action. The studies were based on experimental settings where parents were instructed to teach the functions of ten different objects to their infants. We analysed multi-modal cues from the parents' hand movements on the one hand and the associated speech cues on the other hand when one particular object was presented.
We obtained objective measurements from the parents' hand movements – that can also be used by a robot in a human-robot interaction scenario – by applying automatic tracking algorithms based on 2D and 3D models that were able to track the trajectories of the hand movements based on movies from a monocular camera (Schmidt et al., 2006). A number of variables capturing objectively measurable aspects of the more subjectively defined variables as used by (Brand et al., 2002) were computed. Results confirmed that there are statistically significant differences between child-directed and adult-directed actions. First, there are more pauses in child-directed interaction, indicating a stronger structuring behaviour. Secondly, the roundness of the movements in child-directed action is less than in adult-directed interaction. We define roundness as the covered motion path (in meters) divided by the distance between motion on- and offset (in meters). This means that a round motion trajectory is longer and more common in an adult-adult interaction (Fritsch et al., 2005); similarly to the notion of "punctuation" in (Brand et al., 2002), an action performed towards a child, is less round because it consists of more pauses between single movements, where the movements are shorter and more straight resulting in simpler action chunks. Thirdly, the difference between the velocity in child-directed movements and adult-directed movements shows a strong trend towards significance when measured in 2D. However, measurements based on the 3D algorithms did not provide such a trend. This raises the interesting question whether parents are able to plan their movements by taking into account the perspective of their infant who will mainly perceive the movement in a 2D-plane.
In addition to these vision-based features, we analysed different speech variables derived from the videos. In general, we found a similar pattern as in the movement behaviour (see also (Rohlfing et al., 2006)): Parents made more pauses in relation to their speaking time when addressing their infants than when instructing an adult. However, we observed a significantly higher variance in this verbosity feature between subjects in the adult-adult condition, indicating that there is a stronger influence of personal style when addressing an adult. In more detail, we observed that the beginnings and endings of action and speech segments tend to coincide more often in infant directed interaction. In addition, when coinciding with an action end, the speech end is much stronger prosodically marked in infant directed speech than in adult directed speech. This could be an indication that the semantics of the actions in terms of goals and subgoals are much more stressed when

addressing an infant. Finally, we observed more instances of verbally stressed object referents and more temporal synchrony of verbal stress and "gestural stress", i.e. shaking or moving of the shown object. These findings match previous findings by (Zukow-Goldring, 2006).
From these results, we derived 8 different variables that can be used for (1) deciding whether a teaching behaviour is being shown (2) analysing the structure of the action and (3) assigning meaning to specific parts of the action (see Table1).

Variable	Detecting "when" to imitate	Detecting action end / (sub)goal	Detecting naming of object attribute (colour, place)
Motion roundness	+		
Motion velocity (2D)	+		
Motion pauses		+	
Speech pauses		+	
Coincidence of speech and movement end		+	
Prosodic emphasis of speech end coinciding with movement end		+	
Verbal stress			+
Synchrony of verbal stress and "shaking" movement			+

Table 1. Variables and their functions in analysing a human tutor's behaviour

In order for a robot to make use of these variables, it needs to be equipped with basic interaction capabilities so it is able to detect when a human tutor is interacting with it and when it is not addressed. While this may appear to be a trivial pre-condition for learning, the analysis of the social situation is generally not taken into account (or implicitely assumed) in imitation learning robots. Yet, to avoid that the robot will start to analyse any movements in its vicinity, it needs to be equipped with a basic attention system that enables it to focus its attention on an interaction partner or on a common scene, thus establishing so called joint attention. In the next section, we describe how such an attention system is realized on Barthoc.

4. The Robot Platform Barthoc

Our research is based on a robot that has the capabilities to establish a communication situation and can engage in a meaningful interaction.
We have a child-sized and an adult-sized humanoid robot Barthoc as depicted in Fig. 2 and Fig. 3. It is a torso robot that is able to move its upper body like a sitting human. The adult-sized robot corresponds to an adult person with the size of 75 cm from its waist upwards. The torso is mounted on a 65 cm high chair-like socket, which includes the power supply,

two serial connections to a desktop computer, and a motor for rotations around its main axis. One interface is used for controlling head and neck actuators, while the second one is connected to all components below the neck. The torso of the robot consists of a metal frame with a transparent cover to protect the inner elements. Within the torso all necessary electronics for movement are integrated. In total 41 actuators consisting of DC- and servo motors are used to control the robot. To achieve human-like facial expressions ten degrees of freedom are used in its face to control jaw, mouth angles, eyes, eyebrows and eyelids. The eyes are vertically aligned and horizontally controllable autonomously for object fixations. Each eye contains one FireWire colour video camera with a resolution of 640x480 pixels.

Figure 2. Child-sized Barthoc Junior

Besides facial expressions and eye movements the head can be turned, tilted to its side and slightly shifted forwards and backwards. The set of human-like motion capabilities is completed by two arms, mounted at the sides of the robot. With the help of two five finger hands both deictic gestures and simple grips are realizable. The fingers of each hand have only one bending actuator but are controllable autonomously and made of synthetic material to achieve minimal weight. Besides the neck two shoulder elements are added that can be lifted to simulate shrugging of the shoulders. For speech understanding and the detection of multiple speaker directions two microphones are used, one fixed on each shoulder element. This is a temporary solution. The microphones will be fixed at the ear positions as soon as an improved noise reduction for the head servos is available.

Figure 3. Adult-sized Barthoc

5. System Architecture

For the presented system a three layer architecture (Fritsch et al., 2005) is used consisting of a deliberative, an intermediate, and a reactive layer (see Fig. 4). The top deliberative layer contains the speech processing modules including a dialog system for complex user interaction. In the bottom layer reactive modules capable of adapting to sudden changes in the environment are placed. Since neither the deliberative layer dominates the reactive layer nor the reactive layer dominates the deliberative one, a module called Execution Supervisor (ESV) was developed (Kleinehagenbrock et al., 2004) located in the intermediate layer as well as a knowledge base. The ESV coordinates the different tasks of the individual modules by reconfiguring the parameters of each module. For example, the Actuator Interface for controlling the hardware is configured to receive movement commands from different

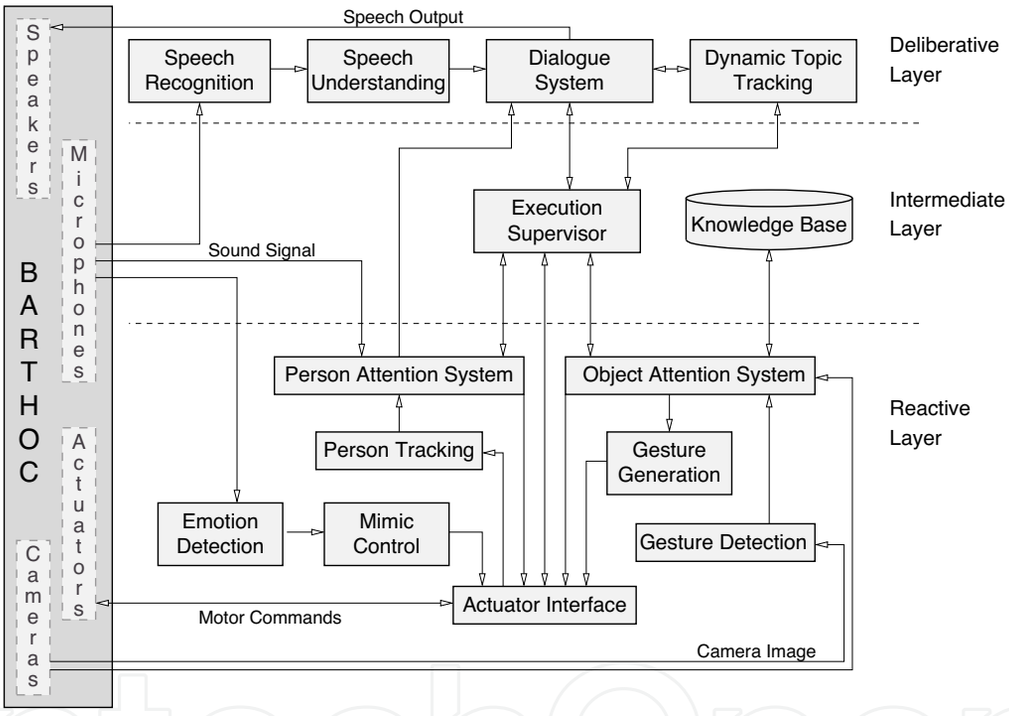


Figure 4. Three layer system architecture of Barthoc, representing the different modules connected by XCF

modules. The ESV can be described as a finite state machine. The different HRI abilities are represented as states and a message sent from a module to the ESV can result in a transition from state A to state B. For each transition the modules in the different layers are reconfigured. Additionally to the new configuration, data like an object label is exchanged between the modules. All data exchange via the ESV is based on the XML Communication Framework (Wrede et al., 2004) using four predefined XML structures, only. All XML data is designed in a human readable style for easy understanding of the internal system communication and efficient debugging.

Using a predefined set of XML structures (see Table 2) data exchange between the ESV and each module is automatically established after reading a configuration file. The file also contains the definition of the finite state machine and the transitions that can be performed. This makes the system easily extendable for new HRI capabilities, by simply changing the configuration file for adding new components without changing one line of source code. Due to the automatic creation of the XML interfaces with a very homogenous structure, fusing the data from the different modules is achieved easily. The system already contains modules for multiple person tracking with attention control (Lang et al., 2003; Fritsch et al., 2004) and an object attention system (Haasch et al., 2005) based on deictic gestures for learning new objects. Additionally an emotion classification based on the intonation of user utterances (Hegel et al., 2006) was added, as well as a Dynamic Topic Tracking (Maas et al., 2006) to follow the content of a dialog. In the next sections we detail how the human-robot interaction is performed by analysing not only system state and visual cues, but spoken language via the dialog system (Li et al., 2006) as well, delivered by the independent operating modules.

```
<MSG xmlns:xs="http://www.w3.org/2001/XMLSchema-instance" xs:type="event">
  <GENERATOR>PTA</GENERATOR>
  <TIMESTAMP>1145869461268</TIMESTAMP>
  <ID>
    <ORIGIN mod="PTA">3</ORIGIN>
  </ID>
  <NAME>CPFound</NAME>
  <STATE>PersonAlertness</STATE>
  <BESTBEFORE>1145869461568</BESTBEFORE>
  <DATA>
    <CPDATA>
      <ID>4</ID>
      <NAME>UNKNOWN</NAME>
  </CPDATA>
  </DATA>
</MSG>
```

```
<MSG xmlns:xs="http://www.w3.org/2001/XMLSchema-instance" xs:type="order">
  <GENERATOR>ESV</GENERATOR>
  <TIMESTAMP>1145870556599</TIMESTAMP>
  <ID>
    <ORIGIN mod="DLG">2</ORIGIN>
  </ID>
  <NAME>FocusCPFace</NAME>
  <STATE>PersonAttention</STATE>
  <DATA>
    <CPDATA>
      <ID>36</ID>
      <NAME>unknown</NAME>
    </CPDATA>
</DATA>
</MSG>
```

Table 2. Examples for data exchange from Person Tracking (PTA) to ESV to inform the system that a person was found (above) and data exchange from Dialog (DLG) via ESV to PTA with the order to focus the face of the current communication partner

6. Finding Someone to Interact with

In the first phase of an interaction, a potential communication partner has to be found and continuously tracked. Additionally, the HRI system has to cope not only with one but also with multiple persons in its surrounding, and thus, discriminating which person is currently attempting to interact with the robot and who is not. The Person Tracking and Attention System is solving both tasks, first finding and continuously tracking multiple persons in the robot's surrounding and secondly deciding to whom the robot will pay attention.

The multiple person tracking is based on the *Anchoring* approach originally introduced by Coradeschi & Saffiotti (2001) and can be described as the connection (Anchoring) between the sensor data (*Percept*) of a real world object and the software representation (*Symbol*) of this object during a fixed time period. To create a robust tracking we extended the tracking from a single to a multi modal approach not tracking a human as one Percept-Symbol relation but as two using a face detector and a voice detector. While the face detector is based on Viola & Jones (2001) the voice detector uses a Cross-Power Spectrum Phase to estimate multiple speaker directions from the signal runtime difference of the two microphones mounted on the robot's shoulders. Each modality (face and voice) is separately anchored and afterwards assigned to a so called *Person Anchor*. A Person Anchor can be initiated by a found face or voice or both if the distance in the real world is below an adjustable threshold. The Person Anchor will be kept and thus a person tracked as long as at least one of its *Component Anchors* (face and voice) is anchored. To avoid anchor losses due to singular misclassifications a Component Anchor will not be lost immediately after a missing Percept for a Symbol. Empirical evaluation showed that a temporal threshold of two seconds increases the robustness of the tracking while maintaining a high flexibility to a changing environment.

As we did not want to restrict the environment to a small interaction area in front of the robot, it is necessary to consider the limited field of view of the video cameras in the eyes of Barthoc. The robot reacts and turns towards people starting to speak outside the current field of view. This possibly results in another person getting out of view due to the robot's movement. To achieve this robot reaction towards real speakers but not towards TV or radio and to avoid loosing track of persons as they get out of view by the robot's movement, we extended the described Anchoring process by a simple but very efficient voice validation and a short term memory (Spexard et al., 2006). For the voice validation we decided to follow the example humans give us. If they encounter an unknown voice out of their field of view humans will possibly have a short look in the corresponding direction evaluating whether the reason for the voice raises their interest or not. If it does, they might change their attention to it, if not they will try to ignore it as long as it persists. Since we have no kind of voice classification any sound will be of the same interest for Barthoc and cause a short turn of its head to the corresponding direction looking for potential communication partners. If the face detection does not find a person there after an adjustable number of trials (lasting on average 2 seconds) although the sound source should be in sight the sound source is marked as not trustworthy. From here on, the robot does not look at it, as long as it persists. Alternatively, a re-evaluation of not trusted sound sources is possible after a given time, but experiments revealed that this is not necessary because the speaker verification is working reliable.

If a person is trusted by the Voice Validation and got out of view due to the robot's movement the short term memory will keep the person's position and return to it later

according to the attention system. If someone gets out of sight because he is walking away the system will not return to the position. When a memorized communication partner re-enters the field of view, because the robot shifts its attention to him it is necessary to add another temporal threshold of three seconds since the camera needs approximately one second to adjust focus and gain for an acceptable image quality. The person remains tracked if a face is detected within this time span, otherwise the corresponding person is forgotten and the robot will not look at his direction again. In this case it is assumed that the person has left while the robot did not pay attention.

The decision to which person Barthoc currently pays attention is taken by current user behaviour as observed by the robot. The system is capable of classifying whether someone is standing still or passing by, it can recognize the current gaze of a person by the face detector and the Voice Anchor provides the information whether a person is speaking. Assuming that people look at the communication partner they are talking to the following hierarchy was implemented: Of the lowest interest are people passing by independently of the remaining information. Persons who are looking at the robot are more interesting than persons looking away. Taking into account that the robot might not see all people as they are out of its field of view a detected voice raises the interest. The most interest is paid to a person who is standing in front of, talking towards and facing the robot. It is assumed that this person wants to start an interaction and the robot will not pay attention to another person as long as these three conditions are fulfilled. Given more than one Person on the same interest level the robot's attention will skip from one person to the next one after an adjustable time span, which is currently four to six seconds. The order for the changes is determined by the order in which the people were first recognized by Barthoc.

Figure 5. Scenario: Interacting with Barthoc in a human-like manner

7. Communicating with Barthoc

When being recognized and tracked by the robot, a human interaction partner is able to use a natural spoken dialog system to communicate with the robot (Li et al., 2006). The dialog model is based on the concept of grounding (Clark, 1992) (Traum, 1994), where dialog contributions are interpreted in terms of adjacency pairs. According to this concept, each interaction starts with a *presentation* which is an account introduced by one participant. This presentation needs to be answered by the interlocutor, indicating that he has understood what has been said. This answer is termed *acceptance*, referring to the pragmatic function it plays in the interaction. Note that the concept of presentation and acceptance does not refer to the semantic content of the utterance. The term *acceptance* can also be applied to a negative answer. However, if the interlocutor did not understand the utterance, regardless of the reason (i.e. acoustically or semantically), his answer will be interpreted as a new presentation which needs to be answered by the speaker, before the original question can be answered. Once an acceptance is given, the semantic content of the two utterances are interpreted as *grounded*, that is, the propositional content of the utterances will be interpreted as true for this conversation and as known to both participants. This framework allows to interpret dialog interactions with respect to their pragmatic function.

Furthermore, the implementation of this dialog model allows to integrate verbal as well as non-verbal contributions. This means, given for example a vision-based head nod recognizer, a head nod would be interpreted as an acceptance. Also, the model can generate non-verbal feedback within this framework which means that instead of giving a verbal answer to a command, the execution of the command itself would serve as the acceptance of the presentation of the command.

With respect to the teaching scenario this dialog model allows us to frame the interaction based on the pragmatic function of verbal and non-verbal actions. Thus, it would be possible for the robot to react to the instructor's actions by non-verbal signals. Also, we can interpret the instructor's actions or sub-actions as separate contributions of the dialog to which the robot can react by giving signals of understanding or non-understanding. This way, we can establish an interaction at a fine grained level. This approach will allow us to test our hypotheses about signals that structure actions into meaningful parts such as sub-goals, means or constraints in an interactive situation by giving acceptance at different parts of the instructor's action demonstration.

8. Outlook

Modelling learning on a robot requires that the robot acts in a social situation. We have therefore integrated a complex interaction framework on our robot Barthoc that it is, thus, able to communicate with humans, more specifically, with tutors. This interaction framework enables the robot (1) to focus its attention on a human communication partner and (2) to interpret the communicative behaviour of its communication partner as communicative acts within a pragmatic framework of grounding.

Based on this interaction framework, we can now integrate our proposed learning mechanism that aims at deriving a semantic understanding of the presented actions. In detail, we can now make use of the above mentioned variables derived from the visual (hand tracking) and acoustic (intonation contour and stress detection) channel in order to chunk demonstrated actions into meaningful parts. This segmentation of the action can be

tested during interactions with human instructors. It will also allow us to analyse the effect of different segmentation strategies, which are reflected in the feedback behaviour of the robot, on the behaviour of the tutor.

9. References

Brand, R.J. ; Baldwin, D. A. & Ashburn, L.A. (2002). Evidence for 'motionese': modifications in mothers' infant-directed action. *Developmental Science* 5, 72-83 (2002).

Caradeschi, S. & Saffiotti, A. (2001). Perceptual anchoring of symbols for action. In : *Proc. Of the 17th IJCAI Conference*, pp. 407-412.

Clark, H.H., ed. (1992). *Arenas of Language Use*. University of Chicago Press.

Fritsch, J. ; Kleinehagenbrock, M. ; Lang, S. ; Fink, G.A. & Sagerer, G. (2004). Audiovisual person tracking with a mobile robot. In *: Proc. Int. Conf. on Intelligent Autonomous Systems.* pp. 898-906.

Fritsch, J. ; Hofemann, N. & Rohlfing K. (2005). Detecting 'when to imitate' in a social context with a human caregiver. In *Proc. IEEE ICRA, Workshop on The Social Mechanisms of Robot Programming by Demonstration.* April, Barcelona, Spain.

Fritsch, J. ; Kleinehagenbrock, M. ; Haasch, A. ; Wrede, S. & Sagerer, G. (2005b). A flexible infrastructure for the development of a robot companion with extensible HRI-capabilitites. In : *Proc. IEEE Int. Conf. On Robotics and Automation*, pp. 3419-3425.

Gergely, G. (2003): What should a robot learn from an infant? Mechanisms of action interpretation and observational learning in infancy. In *Proc. 3d International Workshop on Epigenetic Robotics*. August, Boston, USA.

Haasch, A.; Hofemann, N.; Fritsch, J. & Sagerer, G. (2005). A multi-modal object attention system for a mobile robot. In: *Proc. IEEE/RSJ Int. Conf. On Intelligent Robots and Systems*, pp. 1499-1504.

Hackel, M.; Schwope, M.; Fritsch, J.; Wrede, B. & Sagerer, G. (2006). Designing a sociable humanoid robot for interdisciplinary research. In: *Advanced Robotics*, 20(11): pp. 1219-1235.

Hegel, F.; Spexard, T.; Vogt, T.; Horstmann, G. & Wrede, B. (2006). Playing a different imitation game: Interaction with an empathic android robot. In: *Proc. IEEE-RAS Int. Conf. On Humanoid Robots*, pp. 56-61.

Kleinehagenbrock, M.; Fritsch, J. & Sagerer, G. (2004). Supporting Advanced Interaction Capabilities on a Mobile Robot with a Flexible Control System. In: *Proc. IEEE/RSJ In. Conf. On Intelligent Robots and Systems,* Vol. 3, pp. 3649-3655.

Kuniyoshi, Y., Inaba, M. & Inoue, H. (1994): Learning by watching: Extracting reusable task knowledge from visual observation of human performance. In: *IEEE Transactions on Robotics and Automation*, 10(6), 165-170.

Lang, S. ; Kleinehagenbrock, M. ; Hohenner, S. ; Fritsch, J. ; Fink, G.A. & Sagerer, G. (2003). Providing the basis for human-robot interaction : a multi-modal attention system for a mobile robot. In : Proc. Int. Conf. on Multimodal Interfaces, pp. 28-35.

Li, S. ; Wrede, B. & Sagerer, G. (2006). A computational model of multi-modal grounding. In : *Proc. ACL SIGdial workshop on discourse and dialog, in conjunction with COLING/ACL*, pp. 153-160.

Maas, J.F. ; Spexard, T. ; Fritsch, J. ; Wrede, B. & Sagerer, G. (2006). BIRON, what's the topic ? – A multi-modal topic tracker for improved human-robot interaction. In : *Proc. IEEE Int. Workshop on Robot and Human Interactive Communication,* pp : 26-32.

Rohlfing, K. ; Fritsch, J. ; Wrede, B. & Junmann, T. (2006). How can multimodal cues from child-directed interaction reduce learning complexity in robots, In : *Advanced Robotics*, 20(10) : 1183-1199.

J. Schmidt, J.; Kwolek, B. & Fritsch, J. (2006). Kernel Particle Filter for Real-Time 3D Body Tracking in Monocular Color Images. In: Proc. of Automatic Face and Gesture Recognition, pp. 567-572.

Spexard, T. ; Haasch, A. ; Fritsch, J. & Sagerer, G. (2006). Human-like person tracking with an anthropomorphic robot. In : *Proc. IEEE Int. Conf. on Robotics and Automation*, pp. 1286-1292.

Traum, D. ; Rickel, J. (1994). *A Computational Theory of Grounding in Natural Language Conversation*. Ph.D. Dissertation, University of Rochester.

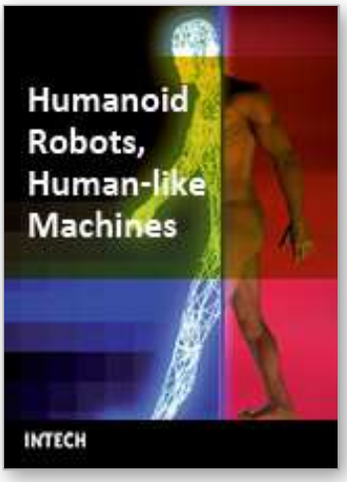
Viola, P. & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *: Proc. Int. Conf. on Computer Vision and Pattern Recognition*, pp. 511-518.

Weber, C., Elshaw, M., Zochios, M. & Wermter, S. (2004). A multimodal hierarchial approach to robot learning by imitation. In L. Berthouze, H. Kozima, C. G. Prince, G. Sandini, G. Stojanov, G. Metta, and C. Balkenius (Eds.): *Proc. 4th International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, pp. 131-134.

Wermter, S., Weber, C., Elshaw, M., Gallese, V. & Pulvermüller, F. (2005). Grounding neural robot language in action. In: Wermter, S., Palm, G. & Elshaw, M. (Eds.): *Biomimetic Neural Learning for Intelligent Robots*, pp. 162-181. Heidelberg: Springer.

Wrede, S. ; Fritsch, J. ; Bauckhage, C. & Sagerer, G. (2004). An XML based framework for cognitive vision architectures. In : *Proc. Int. Conf. on Pattern Recognition*, Vol. 1, pp. 757-760.

Zukow-Goldring, P. (2006). Assisted imitation: Affordances, effectivities, and the mirror system in early language development. In *From action to language*, M. A. Arbib (Ed.). Cambridge: CUP.



Humanoid Robots, Human-like Machines

Edited by Matthias Hackel

ISBN 978-3-902613-07-3

Hard cover, 642 pages

Publisher I-Tech Education and Publishing

Published online 01, June, 2007

Published in print edition June, 2007

In this book the variety of humanoid robotic research can be obtained. This book is divided in four parts: Hardware Development: Components and Systems, Biped Motion: Walking, Running and Self-orientation, Sensing the Environment: Acquisition, Data Processing and Control and Mind Organisation: Learning and Interaction. The first part of the book deals with remarkable hardware developments, whereby complete humanoid robotic systems are as well described as partial solutions. In the second part diverse results around the biped motion of humanoid robots are presented. The autonomous, efficient and adaptive two-legged walking is one of the main challenge in humanoid robotics. The two-legged walking will enable humanoid robots to enter our environment without rearrangement. Developments in the field of visual sensors, data acquisition, processing and control are to be observed in third part of the book. In the fourth part some "mind building" and communication technologies are presented.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen