A Cognitive Architecture for a Humanoid Robot: A First Approach

Catherina Burghart*†, Ralf Mikut*‡, Rainer Stiefelhagen*†, Tamim Asfour*†,
Hartwig Holzapfel*†, Peter Steinhaus*†, Ruediger Dillmann*†
*Collaborative Research Centre on Humanoid Robots - Learning and Cooperating Multimodal Robots (SFB 588)
D-76128 Karlsruhe, Germany
Email: burghart@ira.uka.de, ralf.mikut@iai.fzk.de, {stiefel, asfour, hartwig, steinhaus, dillmann}@ira.uka.de
†University of Karlsruhe, Germany
‡Forschungszentrum Karlsruhe GmbH, Institute of Applied Computer Science, Germany

***Abstract*— Future life pictures humans having intelligent humanoid robotic systems taking part in their everyday life. Thus researchers strive to supply robots with an adequate artificial intelligence in order to achieve a natural and intuitive interaction between human being and robotic system. Within the German Humanoid Project we focus on learning and cooperating multimodal robotic systems. In this paper we present a first cognitive architecture for our humanoid robot: The architecture is a mixture of a hierarchical three-layered form on the one hand and a composition of behaviour-specific modules on the other hand. Perception, learning, planning of actions, motor control, and human-like communication play an important role in the robotic system and are embedded step by step in our architecture.**

Fig. 1. Humanoid robot in the robot kitchen at the SFB 588 lab in Karlsruhe

I. Introduction

There is no denying that robots gradually enter all fields of human everyday life. This leads to challenging research questions like acceptance of a robot by its human user, safety of the humans interacting with the robot, and cognition and artificial intelligence required of the robotic system. Our research is dedicated to the development of concepts, methods and concrete mechatronic components for a humanoid robot sharing its workspace with humans [1]. Key issues of the Collaborative Research Center 588 "Humanoid Robots – Learning and Cooperating Multimodal Robots" (SFB 588) are *multimodal communication channels* allowing the human user to intuitively interact with the robotic system, *cooperation between human and robot* in different contexts, and the robot's *ability to learn* previously unknown tasks, new motions, new concepts and new objects. Our present testbed is a humanoid robot in a kitchen environment (Fig. 1), which comprises a mobile two-arm system with five-fingered hands, a flexible torso, a sensor head with visual and acoustic sensors, and an artificial skin. In addition, the motion system and thus the behaviour of the robot is tailored towards *human-like motions*.

Translating our key issues into a working humanoid robotic system requires an adequate cognitive architecture comprising elementary building blocks for technical cognition and intelligence of the robot. In cognitive psychology cognition includes functions of perception and recognition, of encoding, storing and memorizing, as well as of thinking and problem solving, of motor control, and, finally, of usage of language [2]. Questioning several international researchers in robotics and examining actual publications on intelligent robotic systems showed that actually the researchers apply a similar definition of cognition as used by psychologists to intelligent robotic systems. In contrast to cognitive psychology the different functions of cognition had a different priority. Perception of the robotic system was named first, then learning, motor control, reasoning, problem solving, goal orientation, knowledge representation and communication followed. Self-consciousness, motivation and emotions of a robotic system being functions of cognition were in dispute. Although a similar definition is applied, only some functions are actually realized by the researchers in their robotic systems.

The first cognitive architectures for intelligent robots came from artificial intelligence (see e.g. [3], [4]). Such architectures tried to create a complete symbolic world model of the robot environment using sensor data and to make a plan on a symbolic level (sense-plan-act (SPA) strategies). This strategy has many drawbacks, such as slow reaction times and weak couplings between model representation and the real world scenario. Moravec [3] wrote about the Stanford CART: *The system was reliable for short runs, but slow. The Cart moved one meter every ten to fifteen minutes, in lurches.* Such robots were more or less restricted to live in toy worlds. As a reaction, Brooks proposed a more skill-oriented architecture known as subsumption architecture [5]. This architecture has many parallel interacting components for different functional ”behaviours”. Similar ideas can be found in [6]–[8]. Open problems are the explicit representation of goals and the coordination of the components. A fusion of both approaches has been proposed with slightly different names (i.e. three-layered architecture) from different authors [9]–[11]. Here, only the lowest level contains behaviours. A mid-level coordinates the components of the lowest level and a planning component is placed on the top level. Further developments based on this concept can be found in [12]. Such architectures are related with human operator models from cognitive science subdividing human information processing in (task-related) skill-based, rule-based and planning-based elements [13]. In the last years, different papers proposed the integration of emotions as motivational system [14]–[16]. However, a detailed specification of three-layered architectures for humanoid robots is an open question.

Our notion of cognition of an intelligent robotic system comprises functions of perception, of memorizing and learning, of solving problems (complex task planning), of motor control and of communication (speech, gestures, mimics,...). We have therefore designed a cognitive architecture to meet our needs, which is presented in this paper. The architecture serves as the reference architecture of the German Humanoid Project and is submitted to ongoing efforts and changes.

This paper is organised as follows: Section II gives an overview over our cognitive architecture, whereas perceptual components, dialogue components, and task related components are described in detail in Sections III, IV, and V. Section VI pictures the usage of temporal active models, while learning components are focused by Section VII. The actual state of implementation is described in Section VIII; Section IX concludes this paper.

II. Overall Design of the Cognitive Architecture

A cognitive architecture for an intelligent robotic system is required to support fast perception, control and task execution on a low level as well as recognition and interpretation of complex contexts, planning of intrinsic tasks, and learning of behaviours, which are typically all processed on higher levels. Higher levels correlate with higher complexity and higher understanding. In order to satisfy all these needs we have chosen a three-layered architecture adapted to the requirements of a humanoid robot. It comprises parallel behaviour-based components interacting with each other, which can be triggered by entities on a higher level. The main advantages are fast reaction times to external events, an explicit integration of robot goals in the planning layer and a modular design approach.

The building blocks of the architecture are depicted in Fig. 2. The interface to the robotic environment is formed by the actual robot hardware, that is its sensors and actuators. The single perceptual and task oriented components are distributed on three layers. Fast reactive components (low-level) and building blocks for recognition or task coordination (mid-level) act in real-time. The highest level comprises such challenging components as fusion of different perceptual results and recognition of situations and contexts. A dialogue component, a task planer, and a learning component are also situated on the top-level. A global knowledge database houses several sub-databases required by all components on all layers.

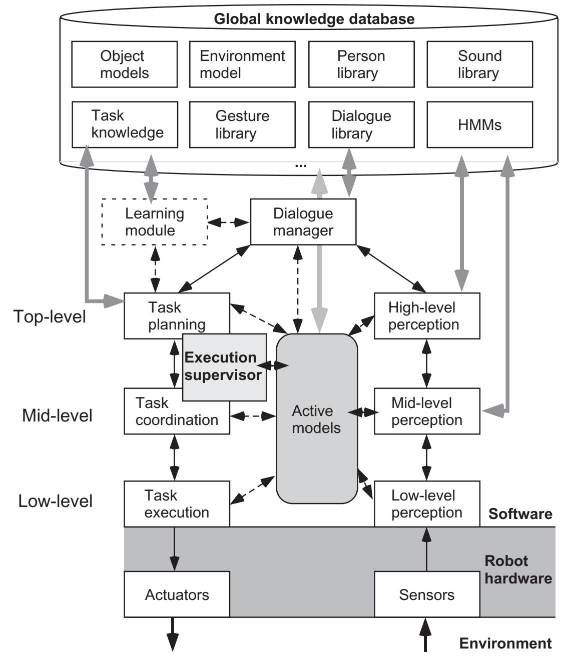

Fig. 2. Cognitive architecture of the Karlsruhe Humanoid Robot

Active models used for real-time perceptions and tasks are retrieved from the global knowledge database and stored in a cache memory. In addition, they have a certain a degree of autonomy for adaptation. The execution supervisor is responsible for the priority management of tasks and perceptual components.

Robot tasks are planned on a higher symbolical level using task knowledge. A resulting sequence of actions is passed on to the task coordination on the mid-level, which then coordinates all running tasks. The final correctly parameterised and deadlock free flow of actions is executed in the low-level task execution.

Interactions between cognitive components across all layers are needed depending on context and behaviour. This concerns the focus of attention of the robotic system, priority management of necessary perceptual components and tasks, access of low-level components to data stored in the knowledge base, dialogue management, or complex reflexes requiring more intrinsic sensor data interpretation than performed by low-level perceptual components.

All components of the architecture are described in detail in the following sections.

III. Perceptual Components

Low-level perception comprises fast interpretation methods of sensor data without any need to access the system knowledge database. These sensor data are relevant for either low-level control of the robotic system and reflexes or as inputs for mid-level perception. The result of the low-level perception is therefore not only passed on to the mid-level perception but also to the task execution via the active models. Especially data coming from joint position sensors, the force torque sensors located in the robot's wrists, data from tactile sensor arrays used as artificial sensitive skin [17], and acoustic data for sound [18] and speech activity detection [19] are processed within this module.

The *middle layer in the perception* hierarchy comprises the various recognition components of the system. Based on the pre-processing results of the low-level perceptual components single modality recognition as well as multimodal recognition (e.g. audio-visual speaker tracking) is possible. These recognition components have access to the database, where persistent information is stored, and also to the active models. The active models (see Fig. 2) serve as a short time memory and provide current environmental information, as well as information about objects in the focus of attention. The list of recognition components should be easily extensible to add new features. Major components are speech recognition [20], [21], acoustic, visual [22] and audiovisual speaker tracking [18], object and person recognition, and gesture [23] recognition.

The *highest level within perception* organizes all understanding components such as single modality understanding, multimodal (late) fusion, and situation recognition. On the highest level of perception the system also interprets actions by the user and recognizes them as communicative or non-communicative behaviour. Modality dependent understanding components are speech understanding, gesture interpretation, movement interpretation [24] and intention prediction. The situation recognition component requires the results of the other components to create a situational representation and interpretation.

Multimodal fusion can take place on various levels from early integration to late integration (for an overview see e.g. [25]). While audio-visual person tracking is classified as a mid-level perceptual component, high-level fusion uses semantic representations as input. Represented as an understanding component, fusion on this level can combine temporally correlated input, such as pointing gestures and speech (e.g. [19]). Even later fusion for communicative input with only loose correlation in time is possible in discourse which is handled by the dialogue manager.

IV. Dialogue Manager

Communicative interaction with the user and interpretation of communicative events is coordinated by the dialogue manager. It serves as an intermediate component between perception and task planning, being able to interpret the pragmatics of user input in the current context, answer questions, request missing information and to forward action requests to the task planner. On the other hand it can be initiated by the system to request information by the user, e.g. to resolve from error states. Spoken or multimodal output needs to be scheduled for execution by the task scheduler, since it requires access to system resources. Communicative output can also influence the attention control system where the system expects new spoken or multimodal input (user needs to be in the field of view) by the user.

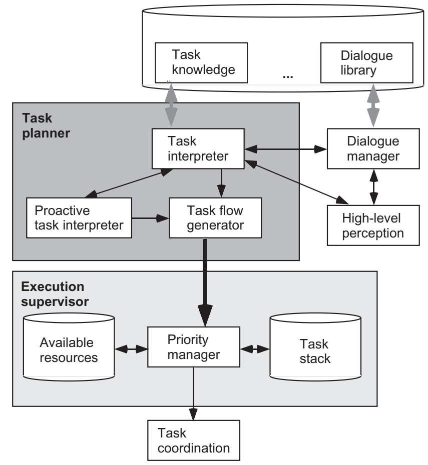

Fig. 3. Components of the task planner and its interaction with the execution supervisor

Furthermore, the dialogue manager has functionality by itself that can be seen as cognitive functions, such as user intention recognition and user modelling. Dialogue strategies should take into account the user's current state such as emotion, system and task conditions and situational constraints, such as higher safety levels [26], [27]. Learning of strategies, new tasks, new formulations and new words, as well as introduction of new objects or persons are attributes that a dialogue system needs to support within the overall cognitive architecture.

V. Task Oriented Components

As a symbolic planner the *task planner* operates above the real-time level using task knowledge. The planning process is started when a desired task has been successfully interpreted out of the data passed from the high-level perception or from the dialog manager. A symbolic plan consists of a sequence of actions which the task flow generator (Fig. 3) selects out of the task knowledge in the knowledge base. Here, a set of different actions with pre- and postconditions is stored in form of XML-files or Petri nets. The actions have been acquired by various learning and programming processes e.g. programming by demonstration or optimisation of actions. The task flow generator assembles the plan for the intended task and adapts the free parameters to the given task. The result of the task planner is a specified plan which can also be divided into several subplans for subsystems e.g. left arm and right arm running in parallel. Additionally, the complete plan includes information, which actions are allowed to be interrupted by other components of the robotic system i.e. if an emergency is detected or if a change of focus of attention is required. The task planner itself can always interrupt actions during runtime and replan them. The task planning is able to delegate decisions to the mid-level by integrating resource management aspects into the plan for task coordination.

The *execution supervisor* is located on the output of the task planner (Fig. 3). It is responsible for the final scheduling of tasks and the resource management of the robot.

Resource management is a hard problem for humanoid robots because many subsystems are involved in different tasks. Resources include hardware subsystems (platform, head, arm(s), hand(s), microphone, speakers) with associated software modules, perceptual software and computer capacities. As an example, the robot head with the cameras is needed for visual servoing, navigation, interaction with the user and localization of external events. Similar problems occur for software components as text-to-speech, object and user recognition.

The proposed concept for resource management and task coordination bases on ideas in [28]. This paper describes a Petri net based strategy and shows an example for three different tasks allocating one, two or three different subsystems as resources. The resources to be allocated are already specified for each task by the task planner. Additionally, resources bound by the attention of the robot are managed by the execution supervisor. The challenge is to manage resource allocation needed for a new focus of attention without any deadlocks.

Petri net based formulation of robotic tasks is quite popular for industrial robots especially for manufacturing tasks (see e.g. [29]). For intelligent robots and machines in general, implementations especially for the coordination layer have been proposed by [30]. But there exist only few papers using Petri nets in humanoid robotics for coordination, hardware resource handling and planning (see e.g. [28], [31], [32]).

The execution supervisor starts the task from the task stack with the highest priority if all necessary hardware and software resources for the task are available to avoid deadlocks. At the same time, all necessary resources are allocated until this task is completed or cancelled. The successful or unsuccessful completion of the task is an integrated message for a defined token configuration of the Petri net in task coordination. In addition, the task planning is able to force the completion of running Petri nets in task coordination (by handshaking) or cancel this net (without handshaking) to free resources for tasks with a higher priority.

With this concept, the execution supervisor dynamically adapts the configuration of the robot to the actual task. A parallel performance of different tasks can be organized by an integrated planning in *one* Petri net or a parallel performance of two or more Petri nets allocating different resources. From a theoretical point of view, the latter case could be interpreted as a composed Petri net with independent subnets.

The *task coordination* is the mid-level component of the three-layered architecture. On this level, a Petri net is executed to coordinate all sub-tasks by firing transitions. The net activates all functions associated with at least one marked place of the Petri net. Examples for such functions are controllers, trajectory planners and reflexes belonging to the lower execution level. In this sense, the task coordination adapts the configuration to the given situation.

An important part of the Petri net is a set of places and transitions for an integrated detection of various exceptions. These transitions process external events, token configurations, timeouts for time durations of tokens at a place and aggregated features from the execution level in a rule-based form. An example for the rule-based detection of instabilities of lower-level control loops is shown in [33]. All rules are associated with places and transitions of the net and stored in the task knowledge. The exception handling is intended as a multi-level functionality starting with an autonomous reaction of the robot to avoid damages followed by a replanning of actions by the robot, asking for user support by the dialogue manager up to a tele-operated exception handling [34].

During a task, the coordination level mostly works quite independently of task planning. The communication is restricted to the starting and stopping of tasks and the handling of unexpected exceptions which can not be solved on the coordination level.

The *task execution* level is characterized by control theory to execute specified sensory-motor control commands. Motion trajectories in the task space are mapped onto robot motor commands. Closed loop control strategies are running in order to meet the desired values. All necessary information about the immediate environment, the current task and the state of the robot are provided by the active models. In this level, mechanical, electrical and sensorial failures are detected and if necessary reported to the task coordination level.

VI. GLOBAL AND ACTIVE MODELS

A cognitive robotic system requires a great variety of models in order to correctly recognise and interpret communication aspects like speech, gestures and mimic, human behaviour, the static environment including objects, and the overall context within which it itself is an actor. These models are all stored in a knowledge base, which is a conglomeration of different sub-databases of the individual model types (i.e. object ontologies and geometries, Hidden Markov Models (HMMs), kinematic models, ...).

Mid- and low-level perception and control need a fast processing of relevant data. This implicates an additional fast access to actual models, sensor data and results. Therefore, our cognitive architecture uses active models as depicted in Fig. 2.

These active models play a central role in this architecture. They are first initialised using global models and are mainly updated by the perceptual system. The novelty of the active models is the ability for their autonomous actualisation and reorganisation. An active model consists of the internal knowledge representation, interfaces, inputs and outputs for information extraction and optionally active parts for update strategies, communication with other active models or global models, learning procedures and logical reasoning.

Data out of the perceptual process are continuously stored and renewed in a cache memory, which can be accessed by all components. The same applies to other information like the current robot state or the actual task. In this way the current task knowledge can be passed on to all components. It is available at the cache memory, as soon as the current task is interpreted and the appropriate actions are selected.

Although all sensor data and perception results are perpetually stored in the active models, there are also closed control loops for reflexes and fast control strategies i.e. zero force control and stop commands. In these cases a direct communication is required, as a fast and adequate reaction of the robotic system cannot be achieved, if the concerning control components permanently have to compete with other modules about the access to the active models.

VII. Learning Components

A prerequisite for an intelligent cognitive system is its ability to learn. Here, different types of learning modes can be distinguished: supervised and unsupervised learning. In both cases the robotic system has to be set into an appropriate learning mode in order to acquire new behaviours and store newly learned facts, tasks and flows of actions in the knowledge base. The learning modes are triggered by the task planner as soon as the task interpreter has correctly recognised the command coming from the dialogue manager or the context coming from the high-level perception. The update and addition of data in the knowledge base is not performed in real-time.

At present, the robotic system learns tasks and flows of actions in an off-line manner by programming by demonstration [35] or by tele-operation [34]. Later, on-line learning procedures like learning by imitation and by tutelage are to be added. Objects, faces, words, phrases, dialogue components are taught to the robotic system in a supervised learning mode.

Additionally, all current states of the robotic system, running times, success statistics and resource allocations of performed tasks are stored in the active models. By comparing the results in the cache memory with the action flow already stored in the task knowledge, the flow can be optimised by the robotic system. In this case the robotic system uses its own performance characteristics in an unsupervised learning mode triggered by the task planner and the execution supervisor.

VIII. Implementation into the System

As outlined before, a number of components of the cognitive architecture have already been implemented and integrated into a working humanoid robot. This section gives a brief overview of our current system.

Our current robot prototype is based on ARMAR [36], a humanoid robot with 23 degrees of freedom. From the kinematics control point of view, the robot consists of five subsystems: Head, left arm, right arm, torso and a mobile platform. The upper body of the robot has been designed to be modular and light-weighted while retaining similar size and proportion as an average person. The control system of the robot is divided into separate modules. Each arm as well as torso, head and mobile platform have its own software- and hardware controller module. The head has 2 DOFs arranged as pan and tilt and is equipped with a stereo camera system and a stereo microphone system. Each of the arms has 7 DOFs and is equipped with 6 DOFs force torque sensors on the wrist. The arms are equipped with anthropomorphic five-fingered hands driven by fluidic actuators [37]. The mobile platform of the robot consists of a differential wheel pair and two passive supporting wheels. It is equipped with front and rear laser scanners and it hosts the power supply and the main part of the computer network.

One of the target scenarios we focus on is a household situation in which a human can ask the robot questions related to the kitchen (such as “What is in the fridge?”), ask the robot to set a table, to switch certain lights on or off, or to bring specific objects. We started by implementing a scenario in which the user asks the robot to bring him a cup. This situation includes different subtasks implemented as different states in a finite state machine:

- localising user
- multimodal interaction with speech and gestures: The user tells the robot which cup it should get from where.
- moving near the cup
- searching cup
- grasping cup
- re-localising user and going back
- delivering cup to the user

Here, we examine different scientific and technical problems, such as multimodal fusion of speech and gestures as in ”get this cup” with a corresponding pointing gesture [20]. Furthermore, errors can happen in every state. For example, the robot did not manage to find the cup: An error message is generated and sent to the dialogue manager which informs the user by speech and waits for an answer to solve the problem. On the other hand, the user might even notice the error before the robot does and alerts it. In both situations, we have to face the problem that the user’s utterances are sometimes elliptical or incomplete and that the user might even shout at the robot in critical situations. In order to integrate such elliptical utterances in the current discourse of the dialogue manager, we set up a context consisting of the current state and also the error message created by the system itself if any. In addition, methods for integrated user localisation by means of acoustic in combination with visual features are explored.

Low-level communication such as camera access and person tracking results is implemented into MCA [38]. High-level communication such as speech recognition results and semantic structures are sent over a loosely coupled agent architecture [20]. Tighter integration with access to the active models is still a future task, as well as the development of missing perceptual components such as situation recognition and a shared situation model.

IX. CONCLUSION

An approach for a cognitive architecture for an intelligent humanoid robotic system has been presented in this paper. Two different paradigms for such an architecture have been combined: on the one hand the architecture follows a strict hierarchical three-layer concept with a low, a middle and a top level. On the other hand behaviour specific components communicating and interacting with each other can be found on all layers. There are cognitive components tailored towards cognitive functions as to be found in human beings. High- and low-level perception play an important role as well as attention control, communication elements, dialogue management, memorizing, learning, complex task planning, and motor control. The architecture design leaves enough room for a future integration of additional components like emotion control, social interaction or new learning modes.

ACKNOWLEDGMENT

This work has been performed within the framework of the German humanoid robotics program (SFB 588) funded by the German Research Foundation (*DFG: Deutsche Forschungsgemeinschaft*).

REFERENCES

[1] R. Dillmann, R. Becher, and P. Steinhaus, "ARMAR II - a learning and cooperative multimodal humanoid robot system," *International Journal of Humanoid Robotics*, vol. 1(1), pp. 143–155, 2004.

[2] G. Strube and K. Wender (Eds.), "The Cognitive Psychology of Knowledge (Advances in Psychology)," *North-Holland*, 1993.

[3] H. Moravec, "Stanford cart and the CMU rover," *Proceedings of the IEEE*, vol. 71(7), pp. 872–884, 1983.

[4] N. J. Nilsson, "Shakey the robot," *SRI International, Tech. Rep. 323*, 1984.

[5] R. A. Brooks, "A robust layered control system for a mobile robot," *IEEE Journal of Robotics and Automation*, vol. 2(1), pp. 14 – 23, 1986.

[6] K. Kawamura, T. Rogers, and X. Ao, "Development of a cognitive model of humans in a multi-agent framework for human-robot interaction," *Autonomous Agents and Multiagent Systems (AAMAS)*, pp. 1379–1386, 2002.

[7] R. Pfeifer and C. Schreier, "Understanding Intelligence," *MIT Press*, 1999.

[8] R. Bischoff, "Design Concept and Realization of the Humanoid Service Robot HERMES," *Field and Service Robotics*, pp. 485–492, 1998.

[9] R. J. Firby, "Building symbolic primitives with continuous control routines," in *Artificial Intelligence Planning Systems*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1992.

[10] J. Connell, "SSS: A hybrid architecture applied to robot navigation," in *Robotics and Automation (ICRA)*, 1992.

[11] E. Gat, "Three-layer architectures." in *Artificial Intelligence and Mobile Robots: Case Studies of Successful Robot Systems*, D. Kortenkamp, Ed. AAAI Press, 1998, pp. 195–210.

[12] M. Hackel, S. Schwope, J. Fritsch, B. Wrede, and G. Sagerer, "A humanoid robot platform suitable for studying embodied interaction," in *Intelligent Robots and Systems*, 2005.

[13] J. Rasmussen, "Skills, rules, knowledge: Signals, signs and symbols and other distinctions in humans performance models," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 13, pp. 257–267, 1983.

[14] C. Breazeal, "Socially intelligent robots: research, development, and applications," *Systems, Man, and Cybernetics*, vol. 4, pp. 2121–2126, 2001.

[15] D. Dörner and K. Hille, "Artificial souls: Motivated emotional robots," in *Systems, Man and Cybernetics*, 1995.

[16] R. C. Arkin, M. Fujita, T. Takagi, and R. Hasegawa, "An ethological and emotional basis for human robot interaction," *Robotics and Autonomous Systems*, vol. 42(3-4), pp. 191–201, 2003.

[17] O. Kerpa, K. Weiss, H. Woern, "Development of a flexible tactile sensor system for a humanoid robot," in *Intelligent Robots and Systems*, 2003.

[18] K. Kroschel, M. Schlosser, and M. Grimm, "Tracking a Moving Person Based on a Microphone Array and a Stereo Camera," 3*DAGA 2005*, Mnchen 2005 .

[19] H. Holzapfel, K. Nickel, and R. Stiefelhagen, "Implementation and Evaluation of a Constraint Based Multimodal Fusion System for Speech and 3D Pointing Gestures," *International Conference on Multimodal Interfaces, (ICMI)*, State College, 2004.

[20] R. Stiefelhagen, C. Fügen, P. Gieselmann, H. Holzapfel, K. Nickel and A. Waibel, "Natural Human-Robot Interaction Using Speech, Head Pose and Gestures, *Intelligent Robots and Systems, IROS*, 2004, Sendai, Japan.

[21] H. Soltau, F. Metze, C. Fügen, and A. Waibel, "A One-Pass Decoder based on Polymorphic Linguistic Context Assignment, In Proc. of the *IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, Madonna di Campiglio, Trento, Italy, December 2001, IEEE.

[22] S. Vacek, S. Knoop, R. Zoellner, and R. Dillmann, "A Framework for Goal-directed, Capability-based Visual Perception, *RO-MAN 2004*, Sept. 20-22, Kurashiki, Japan.

[23] K. Nickel, and R. Stiefelhagen, "Real-time Person Tracking and Pointing Gesture Recognition for Human-Robot Interaction," In: *Computer Vision in Human Computer Interaction*, ECCV Workshop on HCI 2004 Vol. 3058, Prague, Czech Republic, May 2004.

[24] J. Moldenhauer, I. Boesnach, T. Beth, V. Wank, and K. Boes,“ Analysis of Human Motion for Humanoid Robots, *Robotics and Automation ICRA 2005*, 18.22. April 2005, Barcelona.

[25] R. Sharma, V. Pavlovic, and T. Huang, "Toward Multimodal Human-Computer Interface, " *Proceedings of the IEEE*, Vol. 86, No. 5, May 1998.

[26] H. Holzapfel, C. Fügen, M. Denecke, and A. Waibel, "Integrating Emotional Cues into a Framework for Dialogue Management," P*Multimodal Interfaces*, 2002.

[27] H. Holzapfel, and P. Gieselmann, "A Way Out of Dead End Situations in Dialogue Systems for Human-Robot Interaction," *Humanoids*, 2004.

[28] T. Asfour, D. Ly, K. Regenstein, and R. Dillmann, "Coordinated Task Execution for humanoid Robots," *International Symposium on Experimental Robotics 2004 (ISER 04)*, 2004.

[29] D. Milutinovic and P. Lima, "Petri net models of robotic tasks," in *Proceedings Robotics and Automation (ICRA)*, 2002.

[30] F.-Y. Wang and G. N. Saridis, "A coordination theory for intelligent machines." *Automatica*, vol. 26(5), 1991.

[31] K. Kobayashi, A. Nakatani, H. Takahashi, and T. Ushio, "Motion planning for humanoid robots using timed petri net and modular state net," in *Systems, Man and Cybernetics*, 2002.

[32] G. Kim, W. Chung, M. Kim, and C. Lee, "Tripodal schematic design of the control architecture for the service robot PSR," in *Robotics and Automation (ICRA)*, 2003, pp. 2792 – 2797.

[33] R. Mikut, A. Lehmann, and G. Bretthauer, "Fuzzy stability supervision of robot grippers," *Fuzzy Systems*, 2004.

[34] P. Rößler and U. D. Hanebeck,“ Telepresence Techniques for Exception Handling in Household Robots," *Systems, Man and Cybernetics (SMC04)*, 2004.

[35] R. Zoellner, T. Asfour, and R. Dillmann,“ Programming by Demonstration: Dual-Arm Manipulation Tasks for Humanoid, In the IEEE/RSJ *Intelligent Robots and Systems (IROS 2004)*, Sendai, Japan, 2004.

[36] T. Asfour, K. Berns and R. Dillmann, "The Humanoid Robot ARMAR: Design and Control," *Humanoid Robots*, MIT, Boston, USA, 2000.

[37] S. Schulz, Ch. Pylatiuk, A. Kargov, R. Oberle, and G. Bretthauer,“ Progress in the Development of Anthropomorphic Fluidic Hands for a Humanoid Robot," *Humanoid Robots*, Los Angeles, 2004.

[38] T. Asfour, D.N. Ly, K. Regenstein, and R. Dillmann,“ A Modular and Distributed Embedded Control Architecture for Humanoid Robots, *Intelligent Robots and Systems (IROS 2004)*, Sendai, Japan, 2004.