Towards the Development of 'Plug-and-Play' Personal Robots

Rainer Bischoff

Bundeswehr University Munich, The Institute of Measurement Science
85577 Neubiberg, Germany
Bischoff@ieee.org
http://www.unibw-muenchen.de/campus/LRT6

Abstract. Endowing future robot generations with 'plug-and-play' capabilities is one of the fascinating challenges of robotics research. These robots could be easily installed at home and effortlessly used immediately after switching them on. They could be commanded intuitively even by non-experts via multimodal interfaces, would dynamically adapt to ever-changing environmental conditions and never fail. Undoubtedly, current design and safety concepts, locomotion and manipulation capabilities, cooperation and communication abilities, reliability, and – probably most importantly – adaptability, learning capabilities and sensing skills have to be improved significantly to achieve this goal. Therefore, the first part of this paper identifies the major challenges and key technologies on our way towards truly autonomous humanoid robots. It also summarizes and comments current research trends and points out technological gaps that need to be closed by interdisciplinary research efforts. The second part of the paper presents the concepts behind our own humanoid experimental robot *HERMES* that is mainly used to develop techniques for future personal robots. Key areas addressed in this work are human-friendly multimodal man-machine communication and interaction, and vision based mobile manipulation, environmental exploration and navigation. A unifying system architecture has been developed and implemented that integrates state-of-the-art technology under a common software framework. Experiences gained in the development of the robot and in experiments designed to allow an assessment of the overall system performance will be reported.

1 Introduction

How long will it take from now on until robots will be our willing everyday servants that we dream of? Will there ever be robots like R2D2, C3PO or Data whom we know very well from science fiction movies and series and who even develop their own consciousness? If you take a look back at the forecasts that have predicted in recent decades the advent of our robotic "relatives", we have to confess that wishes and reality were (and still are) far apart from each other. Experts announced already in 1964 that we would have domestic robots by the year 1984 and that these robots would be able to autonomously accomplish the major household chores. Even those two-armed mobile helpers that robot pioneer Joseph F. Engelberger predicted in 1993 for the year 1996 will still be for a long time the subject of intensive research and will not be available for most people.

Developing robotic helpers that could be used in many different environments (domestic, public and industrial) for a variety of tasks (e.g., elderly care, helping handicapped people, assistance in factories) is a challenging task. According to [Schraft, Schmierer 1998] the prospects for the necessary key technologies are tremendous, and thus, worth the financial risks and huge research efforts. Undoubtedly, much research is still needed to improve considerably design and safety concepts, locomotion and manipulation capabilities, cooperation and communication abilities, reliability, and – probably most importantly – adaptability, learning capabilities and sensing skills.

Eventually, this research will lead to our ultimate goal – the 'plug-and-play' personal robot – that could be easily bought in any shop, freely configured in hardware and software according to personal needs, plugged in at home (to charge the batteries for the first time), switched on, and from this moment on could be effortlessly operated, i.e., instructed to perform required services. In sharp contrast, most of today's service and industrial robots need to be installed and operated by experts. A 'plug-and-play' personal robot, however, would only need minimal training because most of its skills would be already in-built by the manufacturer. If teaching is required it could be done interactively via a multimodal human-friendly interface that could be intuitively used. To be accepted as a new member of the household or work team the robot would have to act like an intelligent being that is eager to learn more about its new environment by asking people and observing and exploring the environment. It is obvious that such a robot would not be useful if it had to learn everything from scratch and had to be taught by the user like a small child. Instead the robot would require many in-built functions (innate characteristics) that would help it to easily adapt to its new environment.

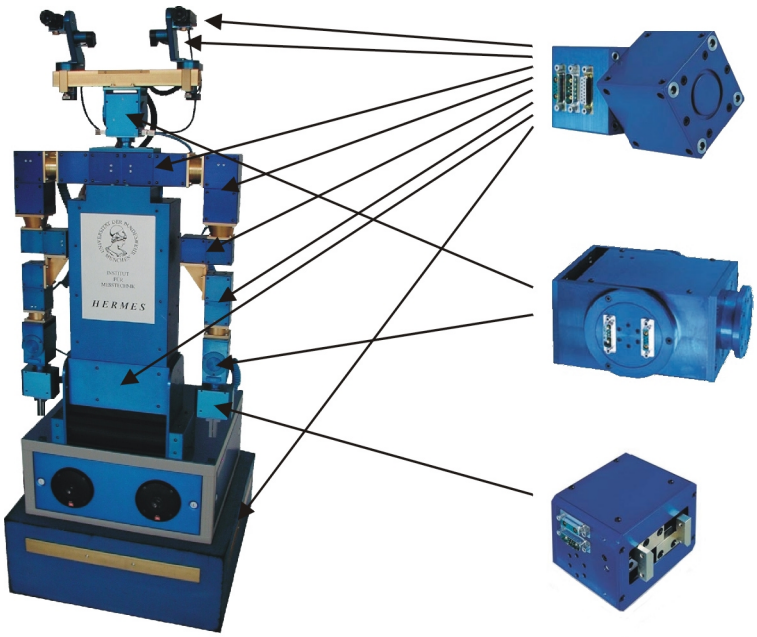

Fig. 1. Humanoid experimental robot *HERMES* built from 25 motor modules with similar mechanical and identical electrical interfaces; each module integrates a motor, a Harmonic Drive gearing, power electronics, a high-resolution angle encoder, a temperature sensor, a current sensor and a microcontroller [Amtec 1996]; *HERMES* consists of 17 rotary modules (top) for propulsion, steering, and arm, body and head movements (head modules with external circuitry box to minimize their moved mass), 3 wrist modules (middle), each with two motor axes for independent pan and tilt movements for the forearms and neck, and two gripper modules (bottom); size of the robot: 1,85 m x 0,70 m x 0,70 m; mass: ca. 250 kg including batteries

To advance research in this promising field and to show what can be done with state-of-the-art components and technology today we have developed the humanoid experimental robot *HERMES*. It is constructed from 25 motor modules with identical mechanical and electrical interfaces, thus yielding a very flexible, extensible and modular design that can be easily modified and maintained (Figure 1). With its omnidirectional wheel base, body, head, eyes and two arms it has now 22 degrees of freedom and resembles a human in height and shape. Its main exteroceptive sensor modality is stereo vision. Both camera "eyes" may be actively and independently controlled in pan and tilt degrees of freedom. A variety of proprioceptive sensors further enhances its perceptual abilities. A multimodal human-friendly interface built upon the basic senses – vision, touch and hearing – enables even non-experts to intuitively control the robot.

The paper consists of two major parts. In the first part, we explain the above-mentioned challenges on our way to the 'plug-and-play' personal robot in greater detail, review the state of the art and try to indicate future research directions that would be suited to close the existing technological gaps (section 2). The second part is devoted to our humanoid experimental robot *HERMES* that in its current state already provides some solutions to the before-mentioned problems. Special attention is given to its design and control architecture (section 3), and some examples of skillful behaviors are described in more detail (section 4). The actual overall system performance can be assessed based on real-world experiments which are described in section 5. Finally, a summary, concluding remarks and an outlook will be given in the sections 6 and 7.

2 Major Challenges for the Development of Personal Robots

From a technical point of view current service robots are intermediate steps towards a much higher goal: "personal robots" that will be as indispensable and ubiquitous as today's personal computers. Personal robots must operate in varying and unstructured environments without needing maintenance or programming. They must cooperate and coexist with humans who are not trained to cooperate with robots and who are not necessarily interested in them. Advanced safety concepts will be as essential as intelligent communication abilities, learning capabilities, and reliability. It is a great challenge to develop these key technologies, but it will be a long way of research to achieve the ultimate goal of a robot, that perceives its environment on a similar level of abstraction as humans, understands complex situations, behaves intelligently and cooperates and communicates with humans

and other machines on a high level of abstraction. In the sequel some of the key technologies will be identified, current research trends will be summarized and commented, and technological gaps that need to be closed by interdisciplinary research efforts will be pointed out.

2.1 Humanoid Robot Design for Human Environments

Personal or service robots shall perform their tasks in environments where humans work and live, e.g., in apartments, offices, laboratories, restaurants, or hospitals. These environments are designed to meet special human characteristics and needs: the space a human requires, his working and vision height and his strength and the number of degrees of freedom available to manipulate objects. Therefore, it seems plausible to take a human as a design model [Bischoff 1997].

Many researchers propose to build a robot-friendly infrastructure, i.e., to install automatic doors, teleoperated lifts and beacons to help keeping the robot structure simple, thereby reducing its costs (e.g., [Kawamura et al. 1996], [Pauly et al. 1999]). This approach might be suitable in the short run for specific system solutions, but only a more generic approach with low-cost universally applicable sensor modalities on the robot will lead in the long run to the deployment of service robots in massive numbers. Instead of adapting the environment to the robot's needs it is a better approach to enable the robot to adapt itself to the environment.

Shaping the robot according to an anthropomorphic model and equipping it with human-like sensor and motor skills will avoid subsequent and expensive changes of the infrastructure and make the robot, in principle, suited for any environment humans normally work and live in. However, for practical reasons it seems plausible to not to use legs, because there are only a few situations where they could be advantageous, e.g., when climbing stairs or stepping over obstacles is required. In most other cases wheels would suffice and be indeed a much better choice because batteries and computers, as well as additional hardware could be easily built into the base without limiting its usability but enormously increasing autonomy.

Another powerful example (apart from *HERMES*) of an autonomous wheeled 'humanoid' robot has been developed by [Bergener et al. 1997]. Their robot ARNOLD consists of a 7-degrees-of-freedom (DoF) arm attached to a TRC Labmate platform. The robot head carries four cameras with different focal lengths (two by two) for executing navigation and manipulation tasks. Several robots exist, mainly in the USA and in Japan, that have some humanoid characteristics; however, it is not possible to describe them all here in full detail. Although most of those systems are immobile or limited in their mobility range due to cables connecting the robots to external power supplies and computers, they possess several DoF in bodies, manipulators and sensor heads. Examples are the robot COG of the MIT AI lab [Brooks, Stein 1994] and the humanoids of Waseda University (WABOT, HADALY, WABIAN and WENDY, see, e.g., [Morita et al. 1999]), Tokyo University (Saika and its successors H3, H4 and H5, see, e.g., [Konno et al. 1997]) and ATR Human Information Processing Research Laboratories (Dynamic Brain DB, see, e.g., [Atkeson et. al. 2000]). While at Waseda and Tokyo University the development of powerful mechatronic components for the humanoids has been in the center of interest, the COG and DB projects aim more at the development of a cognitive system that develops and acts similarly to a person.

From a mechatronic point of view the most sophisticated humanoids up to date are the walking robots P2 and P3 of Honda Motor Corporation. They resemble closely a human in height, shape, and configuration of its degrees of freedom. They are able to balance themselves automatically if pushed, keep themselves upright even on a slope, and climb stairs or slopes. These characteristics enable the robots to perform service tasks such as pushing a cart and tightening bolts [Hirai et al. 1998]. A new initiative is on its way to let these robots play soccer and to teleoperate them through sophisticated interfaces [Honda 1999]. However, until now, they can only walk autonomously (for about 20 minutes) and have very limited environmental perception capabilities.

2.2 Combining Locomotion and Manipulation

Constituting research areas in themselves, robot locomotion and manipulation have for a long time been treated rather independently. Powerful navigation algorithms have been developed in recent years, e.g., [Thrun et al. 1998]. With the advent of powerful mechatronic devices design and control of (industrial) robotic manipulators has been improved considerably, e.g. [DLR 2000]. The integration of both functionalities into complete systems, mobile manipulators, came only recently into the center of interest. A mobile platform yields a significantly enlarged work space compared to a fixed manipulator. The degrees of freedom of the system are increased and redundancy could be used, e.g., to open doors [Nagatani, Yuta 1998]. Obviously, the control of such a manipulator becomes more complex. Major problems to be solved are the uncertainty in the modeling of a mobile manipulation system and to resolve the inherent redundancy of locomotion/manipulation degrees of freedom and of sensor information.

All proposed solutions for this problem have in common that they need tediously calibrated sensors and actuators, world models and knowledge of the kinematic structures. Examples of impressive robots that have been realized based on these principles are, e.g., the assembly robot KAMRO [Lueth et al. 1995], the service robot ROMAN with a remarkable number of key components for service robots [Hanebeck et al. 1997] and the rehabilitation robot MOVAID [Dario et al. 1995]. [Yamamoto 1994] and [Khatib et al. 1995] have developed force control algorithms for multiple cooperating mobile manipulators. [Cameron et al. 1993] developed a mobile manipulator that moved its arm into a favorable manipulation position during the docking phase by reactive control methods.

Given personal or service robotics application scenarios, it is simply impractical to have a fully calibrated system and to employ classical control methods. [Graefe, Maryniak 1998] introduced a promising approach by employing the "sensor-control Jacobian" such that this serious disadvantage can be avoided, with a certain potential to extensions to any desired number of degrees of freedom. The next challenge is to provide a calibration-free solution in combination with a method to resolve redundancy. Problems related to redundant kinematics have not yet been treated in the context of calibration-free robots, and proposed solutions (see above) are still based on careful environmental and kinematic modeling.

2.3 Perception

Model-based robot control depends on a continuous flow of numerical values describing the current state of the robot and its environment. These values are derived from measurements performed by means of the robot's sensors. One problem here is that the quantities needed for updating the numerical models may be difficult to measure, e.g., the distance, mass and velocity of some external object that is posing a collision danger. Also, there are certain important decisions that cannot be made on the basis of measurements alone; the hypothetical decision whether in a particular situation a collision with a parked car should be brought about in order to avoid a collision with a pedestrian is an example.

Humans and other organisms, on the other hand, do not depend on measurements for controlling their motions. If, for instance, we want to sit down on a chair or pass through an open door, we do not first measure the size of the chair, the door, or our body; rather, we make a qualitative judgement whether the chair is high or low, or whether the door is wide or narrow, and then execute a sequence of motions that is adequate for the situation. In short, we substitute perception for measurement.

According to Webster's Dictionary [Babcock 1976] "perception" is a result of perceiving, a reaction to sensory stimulus, direct or intuitive recognition, the integration of sensory impressions of events in the external world by a conscious organism, the awareness of the elements of the environment. "To perceive" means, according to the same source, to become aware of something through the senses, to become conscious of something, to create a mental image, to recognize or identify something, especially as a basis for, or as recognized by, action. Typical questions to be answered by perception are: Which objects exist? What is the relationship between objects? Is it necessary to react? How? Perception, rather than measurement, is thus a prerequisite for, and a complement of, situation assessment.

Vision is the ideal sensing modality for perceiving the environment because it is capable of supplying very rich information on the environment, e.g., to locate a path to be traversed and to judge its apparent condition or to classify objects and to estimate their state of motion. Vision is, therefore, the preferred sensor modality for motion control of higher animals. With the exception of those species adapted to living in very dark environments, they use vision as the main sensing modality for controlling their motions. Observing animals, for instance, when they are pursuing prey or trying to escape a predator, may give an impression of the potential of organic vision systems for motion control.

Employing special purpose sensors, such as sonar, laser or radar, for distance measurement, may help in the short run with the development of specific systems; however, a more generic approach with sensors on the basis of vision, touch and hearing will in the long run lead to more robust and cost-effective solutions. When comparing the most highly developed organic sonar systems with organic vision systems, it is obvious that in all environments where vision is physically possible animals endowed with a sense of vision have, in the course of evolution, prevailed over those that depend on sonar. This may be taken as an indication that vision has, in principle, a greater potential for sensing the environment than sonar. Likewise, it may be expected that advanced robots of the future will also rely primarily on vision for perceiving their environment, unless they are intended to operate in other environments, e.g., under water, where vision is not feasible or ineffective.

One apparent difficulty in implementing vision as a sensor modality for robots is the huge amount of data generated by a video camera: about 10 million pixels per second, depending on the video system used. Nevertheless, it has been shown (e.g., by [Graefe 1989]) that modest computational resources are sufficient for realizing real-time vision systems if a suitable system architecture is implemented. As a key idea for the design of

efficient robot vision systems the concept of object-oriented vision was proposed. It is based on the observation that both the knowledge representation and the data fusion processes in a vision system may be structured according to the visible and relevant external objects in the environment of the robot. For each object that is relevant for the operation of the robot at a particular moment the system has one separate "object process."

2.4 Human Modes of Communication

A user-friendly interface is a prime prerequisite for personal robots that are aimed to help us in various activities in daily life. Many researchers are working towards the goal of truly human-friendly robots that have a number of different senses and can be safely operated and intuitively instructed. Although vision, touch and natural language processing are major components in realizing human-friendly robotics interfaces they have been studied rather independently because they constitute research areas in themselves. Therefore, hardly any work on real robots operating in real environments and integrating all three components has been reported.

Since natural language is the easiest and most desirable mode of communication for a human it is desirable to integrate speech recognition and speech output to most service robots. Acoustically transferred utterances reach the robot even if it is not sight (but around the corner) or within the reach of hands. To equip the robot with a sense of hearing, noise reduction algorithms have to be improved significantly, multi-microphone arrays may have to be used, and speaker-independent speech recognition has to be further enhanced. To understand instructions in natural language the robot must be able to interpret them in a context-dependent way because human conversation is mostly situated, i.e., humans often refer to past and current utterances and to perceived environmental features. Gesture recognition and interpretation, as well as tactile sensor information are needed to fully support a truly human- friendly interface.

[Laengle et al. 1995] have developed a natural language interface for specifying assembly tasks for the dual-arm mobile robot KAMRO. Spatial relations between components to be assembled on the robot's workbench are identified by an overhead camera. Thus, a complete world model can be maintained and the operator's utterances, which may be incomplete or inaccurate, can be complemented or corrected with the help of the world model.

The collaborative research project "Situated Artificial Communicators" (SFB 360) funded by the German Research Organization DFG aims at the discovery of linguistic and cognitive characteristics of human intelligence for communication purposes. The primary testbed for the techniques developed is an assembly cell consisting of two cooperative robot arms equipped with multiple video cameras and force/torque sensors [Knoll et al. 1996], [Zhang et al. 1998]. Typed natural language is used to conduct the conversation between the machine and a human operator. Based upon a simulation of this environment [Milde et al. 1997] have developed a hybrid system architecture that integrates language, perception and action, thereby handling situated references to actions and objects which cannot be interpreted on the basis of the dialogue context alone. Language is treated like any other sensor data and integrated through a behavior-based system.

[Thórisson 1999] created a prototype communicative humanoid, Gandalf, capable of real-time face-to-face dialogue with human users. In the current implementation, Gandalf is a voice, a hand and a face which appear on a small monitor in front of the user. Gandalf is implemented in an architecture called Ymir, a computational model of human psychosocial dialogue skills. The interaction between Gandalf and a human is truly multimodal, i.e., Gandalf produces multimodal behaviors (e.g., various hand gestures, facial expressions, body language and meaningful utterances) and coordinates them based on perception and interpretation of user behavior with respect to the current dialogue state (e.g., gesture recognition, eye tracking, prosody and speech content, direction of head and body, positions of hands in space).

2.5 Situated Reference Frame

In general, robots do not have the perceptual abilities of humans and, therefore, might not be able to detect the features of the environment a human would like to refer to during communication. In other words: While a human might want to describe his world in terms of salient features that he perceives or knows about, a robot has to use certain sensor readings and build special representations that, in general, do not conform with the way humans communicate and interact with each other. However, communication can nevertheless be made possible by virtue of using common reference points in the environment and employing a common labeling for these points. For instance, language could be used to agree upon place or object names as a basis for cooperation [Graefe, Bischoff 1997]. This does not necessarily require an understanding of the robot's representations of the world as long as the robot is able to refer to those agreed-upon points in its own terms when needed.

[Torrance 1994] presented a natural language interface for a mobile robot navigating in an indoor office environment. It is based on predefined (written) instructions that enable a user to supervise the robot during navigation and to provide arbitrary descriptions for specific locations during the map generation process, thus building a basis for a human-friendly reference to places names. The robot can be instructed to navigate according to this common understanding of place names and can give information about its current status and the relationship between places upon request. [Matsui et al. 1997] have developed the talking mobile robot Jijo-2 that is able to build a probabilistic map of its office environment by acquiring missing location information through conversational dialogues with people. A multi-microphone array has been developed to enhance the speech recognition rate under noisy real-world conditions. Speech recognition is based on predefined phonetic dictionaries and syntax tables, and is, thus, speaker-independent. An event-driven multi-agent architecture has been chosen to control the robot's navigation behavior and to conduct dialogues.

2.6 Safety Concept

Safety is a major concern for everybody working with or developing and manufacturing industrial robots. Nevertheless, this seems to be neglected by most robotic researchers. However, a safety concept is needed that early recognizes and reliably avoids dangerous situations. Such a concept has to be different from classical safety concepts for industrial robot settings where the work spaces of operator and robot are strictly separated. In sharp contrast, most service scenarios depend on a close interaction of operator and robot. Therefore, safety needs to be sufficiently enhanced by employing either slip clutches in the joints of manipulators or by implementing intelligent control algorithms that continuously predict and verify force and torque on all joints. Prerequisite for the latter safety concept would be a lightweight manipulator such as presented in [DLR 2000] that allows position, velocity and torque control.

Another (or a complementing) solution could be to mount tactile sensors onto the manipulator's surfaces. For example, [Hoshino et al. 1998] are working on a whole-body tactile sensor suit that can cover arbitrarily shaped robots. Although the suit is not fixed to the robot's body, it can nevertheless provide valuable contact information through dynamically changing mappings of the tactile data on the robot's 3-D surface by compensating self-interference during the robot's own motion. External interaction with a user has been used to safely adapt a 36 DOF humanoid robot to a human's shape while holding it.

In addition, generating smooth trajectories for the robot's joints to yield human-like motion patterns would help to increase overall safety of the robot because such movements can more easily be predicted even by humans who are not interested in robot technology, besides the fact that such fluent movements are emotionally more pleasant.

2.7 Adaptability and Learning

Adaptability and learning facilitate the actual deployment of a robot in a new working environment and enable it to flexibly react to environmental changes and to learn from both successful and erroneous behavior. Compensation of partial sensor failures [Ward, Zelinsky 1997] is desired as well as an adaptation to parameter changes due to maintenance work or aging [Graefe 1995]. A breakthrough in the robotics market will be only made possible, if future personal robots can adapt autonomously to new conditions, because the delegation of highly qualified technicians or engineers for installation and explicit programming of the personal robots would be too expensive.

The problem area of adaptability and learning is closely related to the question to what extent robots should be equipped with innate skills. It is not useful to let a robot learn its basic behaviors or skills from scratch. On the other hand, if it could be done with a prototype of a family of robots, how could the automatically obtained knowledge, the experiences and skills be transferred to its family members, successors and relatives?. We argue that in the first place – for the time being – the robot designer has to take care of what the robot should learn and how. This seems to be the only way to ensure a fail-safe operation and portability of databases and skills.

After switching on a personal robot for the first time it should have access to a multitude of in-built skills that will enable it to learn more about its environment and the people living therein. Potential users of the robot should be able to teach the robot missing parameters in order to personalize it for the new environments and the tasks assigned. Only pre-defined task parameters may be modified during this kind of adaptation process, e.g., maps may be built by the robot and user-specific names may be given to task locations and objects to be manipulated. The robot should be eager to learn more and more things about its environment, but only in a designer-specified way. The drawback of this approach is that robots will only be able to recognize and handle predefined situations which, nevertheless, may include error handling capabilities and would suffice to characterize such a robot as useful and intelligent.

2.8 Robot System Architecture and Flexible Software Framework

A robot system architecture describes a set of architectural components and how they interact. The more complex the system the greater the need for a consistent software framework, i.e. an 'operating system', that integrates all architectural components and handles communication and synchronization among them. This holds especially true when the robot's software development is distributed among many people, and hardware components have to be added, removed or exchanged from time to time. Since all present "intelligent" robots are more or less working prototypes it is not surprising at all that each robot has its own software framework. Somehow, it is a fascinating trait of robotics that there is such a great variety of software frameworks that are closely linked to a manifold of system architectures. However, we should admit that many systems are built in a less structured way, and therefore, cannot be easily modified or ported to even similar hardware. The reason for this is that integration work is regarded as having little or no scientific value, and thus, is frequently neglected. However, such a framework would allow robotics research to be more consistent, i.e. not to invent basic system capabilities again and again and enable skill transfers from one robot to another in spite of different hardware configurations. To cope with these problems [Schlegel, Wörz 1999] developed the software framework SmartSoft that implements complex sensorimotor systems in a modular way by employing object-oriented programming techniques. It has proved its usefulness within the German collaborative research project SFB 527.

Despite these too often neglected 'implementation details', robot system architectures are not yet ready to cope with full-scale humanoid robots. Solutions could be found in nature where biology proves that even the simplest creatures are capable of intelligent behavior: They survive in the real world and compete or cooperate successfully with other beings. Why should it not be possible to endow robots with such an intelligence? By studying animal behavior, particularly their underlying neuroscientific, psychological and ethological concepts, robotic researchers have been enabled to build intelligent behavior-based robots according to the following principles [Bischoff, Graefe 1998]:

- complex behaviors are combinations of simple ones, complex actions emerge from interacting with the real world
- behaviors are selected by arbitration or fusion mechanisms from a repertoire of (competing) behaviors
- behaviors should be tuned to fit the requirements of a particular environment and task
- perception should be actively controlled according to the actual situation

In recent years this type of system architectures has become very popular in robotics research. Many different 'incarnations' have been presented that have more or less proved their effectiveness (for a good overview see [Arkin 1998]). As one realization out of this class of architectures the concept of situation-oriented behavior-based control has been proposed [Bischoff et al. 1996]. Its main characteristics are active perception of the robot's dynamically changing environment, recognition and evaluation of its current situation, and dynamic selection of behaviors appropriate for the actual situation. Animals' simplest capabilities, i.e., to perceive and act within an environment in a meaningful and purposive manner, can thus be imitated by the robots to a certain degree.

2.9 Summary

By reviewing this non-exhaustive collection of challenging design features and problem solutions it becomes obvious that we already possess a large fund of key components for humanoid robots. But how could they be integrated efficiently into a single autonomous and self-contained system, independent of external power lines and computers? To achieve human-level intelligence is certainly a too ambitious goal for the current state of the art. However, the time is ripe for pursuing a unifying approach that integrates the before-mentioned core technologies in order to advance humanoid robotics research on a higher level of abstraction. Only building useful full-scale humanoid robots and putting them into (admittedly limited) personal service of non-experts under real-world conditions will enable us to gather further experiences and to find out what the missing pieces are towards the truly autonomous 'plug-and-play' personal robot.

3 The Humanoid Experimental Robot *HERMES*

The main motivation behind the *HERMES* project is to develop some of the key technologies for future robotic systems and to integrate them in a single autonomous and self-contained testbed, and to evaluate the whole system under real-world conditions.

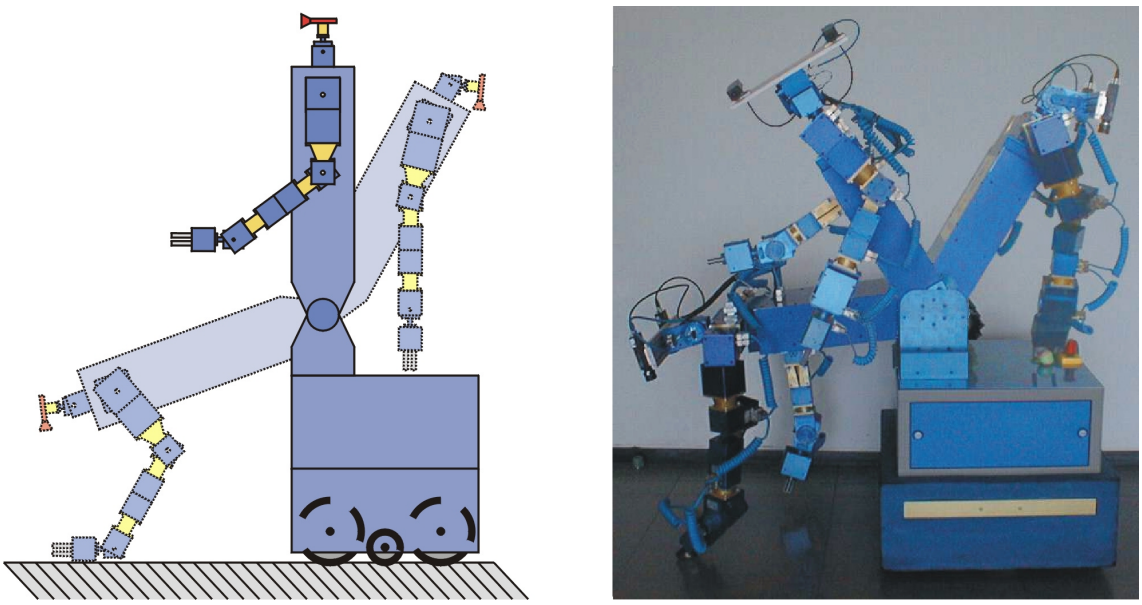

Fig. 2. Left: preliminary design study (1996) to demonstrate the significantly enlarged work space gained by a bendable body. The sensor head remains always in a favorable position for observing end effector activities; Right: Motion sequence of the real robot (1998); the robot's camera head then had two DoF (now six); the two arms have 6 DoF each and a two-finger gripper; one motor module bends the body forward (max. angle: 130°) and backward (max. 90°)

3.1 Design and Realization of *HERMES*

In designing our humanoid experimental robot we placed great emphasis on modularity and extensibility [Bischoff 1997]. All drives are realized as modules with compatible mechanical and electrical interfaces; each drive module consists of two cubes rotating relative to each other and containing a motor-transmission combination, power electronics, sensors, a micro-controller, and a communication interface. A standardized CAN bus connects all drive modules with the main computer. *HERMES* runs on 4 wheels, arranged on the centers of the sides of its base. The front and rear wheels are driven and actively steered, the lateral wheels are passive.

The manipulator system consists of two articulated arms with 6 degrees of freedom each on a body that can bend forward (130°) and backward (90°). The work space extends up to 120 cm in front of the robot. The heavy base guarantees that the robot will not loose its balance even when the body and the arms are fully extended to the front. Currently each arm is equipped with a two-finger gripper that is sufficient for basic manipulation experiments (Figure 2).

Main sensors are two video cameras ("eyes") mounted on independent pan/tilt drive units in addition to the pan/tilt unit ("neck") that controls the common "head" platform. The cameras can be moved very fast with accelerations up to 10.000 °/s^2 and velocities up to 800 °/s. The common pan/tilt unit achieves accelerations of 860 °/s^2 and velocities of 215 °/s. Numerous proprioceptors, such as angle encoders, current converters and temperature sensors, are integrated in the motor modules; additional sensors may be connected via available interfaces. A radio Ethernet interface allows to communicate with the robot from a distance. A wireless keyboard may be used to teleoperate the robot up to distances of 7 m. Separate batteries for the motors and the information processing system allow a continuous operation of the robot for several hours without recharging.

3.2 Behavior-Based Control of a Humanoid Robot

Seamless integration of many – partly redundant – degrees of freedom and various sensor modalities in a complex robot calls for a unifying approach. We have developed a system architecture that allows integration of multiple sensor modalities and numerous actuators, as well as knowledge bases and a human-friendly interface. In its core, the system is behavior-based, which is now generally accepted as an efficient basis for autonomous robots [Arkin 1998]. However, to be able to select behaviors intelligently and to pursue long-term goals in addition to purely reactive behaviors, we have introduced a situation-oriented deliberative component that is responsible for situation assessment and behavior selection.

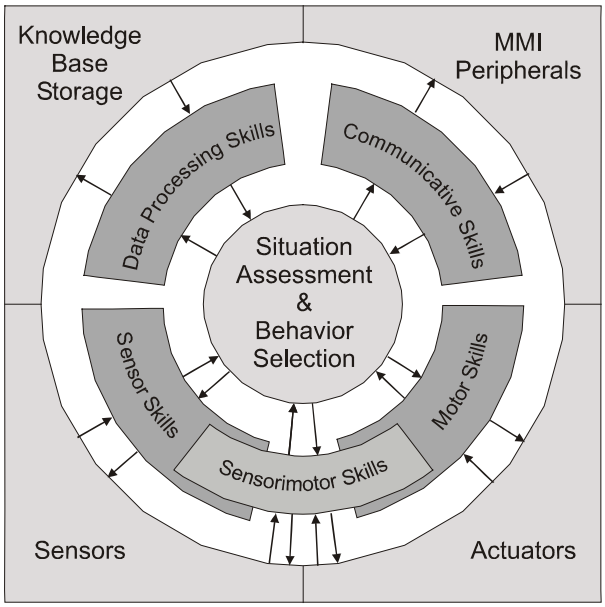

Fig. 3. System architecture of a personal robot based on the concepts of situation, behavior and skills

System Overview. Figure 3 shows the essence of the situation-oriented behavior-based robot architecture as we have implemented it. The situation module (situation assessment & behavior selection) acts as the core of the whole system and is interfaced via "skills" in a bidirectional way with all other hardware components – sensors, actuators, knowledge base storage and MMI peripherals (man-machine and machine-machine interface peripherals).

These skills have direct access to the hardware components and, thus, actually realize behavior primitives. They obtain certain information, e.g., sensor readings, generate specific outputs, e.g., arm movements or speech, or plan a route based on map knowledge. Skills report to the situation module via events and messages on a cyclic or interruptive basis to enable a continuous and timely situation update and error handling.

The situation module fuses via skills data and information from all system components to make situation assessment and behavior selection possible. Moreover, it provides general system management (cognitive skills). Therefore, it is responsible for planning an appropriate behavior sequence to reach a given goal, i.e., it has to coordinate and initialize the in-built skills. By activating and deactivating skills, a management process within the situation module realizes the situation-dependent concatenation of elementary skills that lead to complex and elaborate robot behavior.

In general, most skills involve the entire information processing system. However, at a gross level, they can be classified into five categories besides the cognitive skills: Motor skills are simple movements of the robot's actuators. They can be arbitrarily combined to yield a basis for more complex control commands. Encapsulating the access to groups of actuators, that form robot parts, such as wheelbase, arms, body and head, leads to a simple interface structure, and allows an easy generation of pre-programmed motion patterns. Sensor skills encapsulate the access to one or more sensors, and provide the situation module with proprioceptive or exteroceptive data. Sensorimotor skills combine both sensor and motor skills to yield sensor-guided robot motions, e.g., vision-guided or tactile and force/torque-guided motion skills. Communicative skills pre-process user input and generate a valuable feedback for the user according to the current situation and the given application scenario. The system's knowledge bases are organized and accessed via data processing skills. They return specific information upon request and add newly gained knowledge (e.g., map attributes) to the robot's data bases, or provide means of more complex data processing, e.g., path planning. For a more profound theoretical discussion on our system architecture which bases upon the concepts of situation, behavior and skill see [Graefe, Bischoff 1997] and [Bischoff, Graefe 1999].

Implementation. A hierarchical multi-processor system is used for information processing and robot control. The control and monitoring of the individual drive modules is performed by the sensors and controllers embedded in each module. The main computer is a network of digital signal processors (DSP, TMS 320C40) embedded in a standard industrial PC. Sensor data processing (including vision), situation recognition, behavior selec-

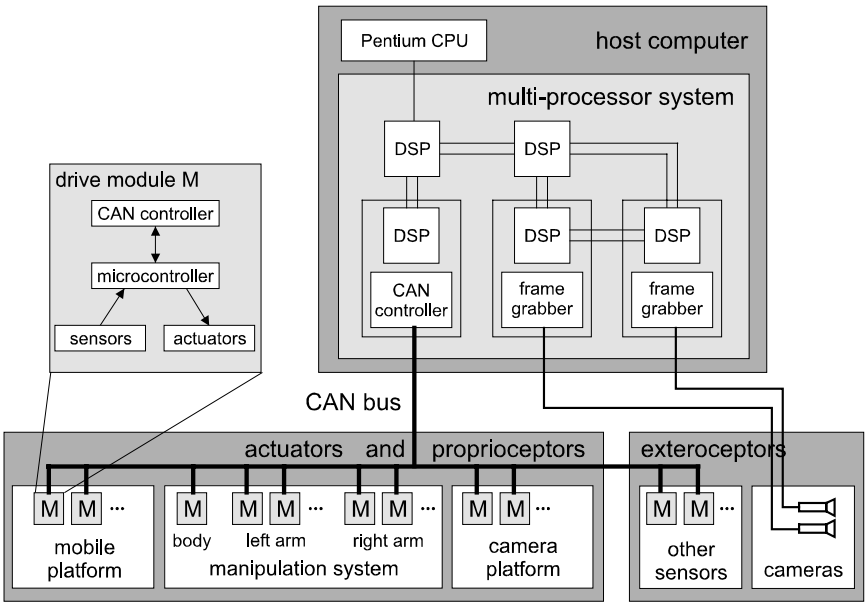

Fig. 4. Modular and adaptable hardware architecture for information processing and robot control

tion and high-level motion control are performed by the DSPs, while the PC provides data storage and the human interface (Figure 4).

A robot operating system has been developed that allows sending and receiving messages via different channels among the different processors and microcontrollers. All tasks and threads run asynchronously, but can be synchronized via messages or events. The left and the right cameras are electrically synchronized, but software needs to make sure that pairs of images taken at the same time are used for stereo vision. However, image acquisition can as well run asynchronously, e.g., using two different image capture rates for two independent image processing tasks.

Overall control is realized as a finite state machine capable of responding to prioritized interrupts and messages. After powering up, the robot finds itself in the state "Waiting for next mission description". A mission description is provided as a text file that may be either loaded from a disk or received via e-mail or entered via the keyboard. It consists of an arbitrary number of single commands or embedded mission descriptions that let the robot perform a required task. All commands are written in natural language and passed to a parser and an interpreter. If a command cannot be understood, or is under-specified or ambiguous, the situation module tries to complement missing information from its situated knowledge or initiates a dialogue with the users to provide it.

In the current implementation commands may be typed, sent via e-mail or spoken, and the robot's responses are written to a display, sent back via e-mail or spoken as well. Motion skills are mostly implemented at the micro controller level within the actuator modules. High-level motor skills, such as coordinated smooth arm movements are realized by a dedicated DSP interfaced to the microcontrollers via a CAN bus. Sensor skills are implemented on those DSPs that have direct access to digitized sensor data, especially digitized images.

4 Realization of Skillful Behavior

In this section we show how skillful goal-directed behavior can be achieved by combining simple motor and sensing skills to more complex sensorimotor skills, and describe how communicative and data processing skills are realized.

4.1 Motor Skills

Motor skills are simple, yet fundamental, movements of the robot's joints, e.g., moving a single module to a certain position or changing its current velocity. High-level motor skills provide access to groups of modules that form specific robot parts (e.g., wheelbase, arms, or head), and generate more complex motion patterns, e.g., they move the arms to a certain position relative to their actual position or set new velocities for all the modules at the

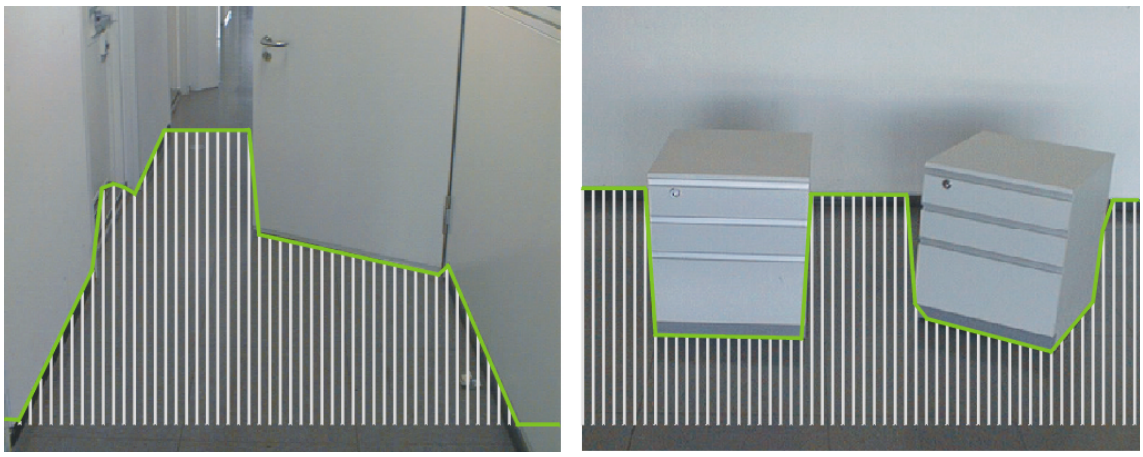

Fig. 5. Gradient filtering along vertical search paths yields contour points that mark the transition between the floor and other objects. Left: typical corridor image with an open door as a possible obstacle. Right: two tables as possible docking objects

same time. Moving an arm requires the definition of ramp parameters (end position, maximum velocity and acceleration) for each joint to reach a given end position. To generate smooth arm movements at each positioning request, it is furthermore required that all modules start and finish their movements at the same time. Therefore, the ramp parameters are individually computed for each module, considering the motion capabilities of the slowest one at a given moment. The governing DSP finally transmits the parameters to the microcontrollers that actually provide accurate ramp control at a rate of 1 kHz.

4.2 Sensor Skills

Sensor skills encapsulate in their simplest form access to the proprioceptive (joint angles, motor currents, battery voltage, etc.) or exteroceptive sensor data (visual, tactile, hearing, etc.).

Visual sensing. One of the most needed sensor skills is to detect objects in the robot's surroundings. Among the objects that a mobile robot needs to detect while navigating in a building are corridors, junctions, doors, work places (e.g., tables) and information signs (e.g., door plates). Obstacles need to be detected as well, but since they may have arbitrary appearance in terms of shape, texture and rigidity, a method that would be suitable for the detection of all kinds of obstacles cannot be given. Instead, we make some basic assumptions about the appearance of the background when it is obstacle-free (see, e.g., [Horswill 1994]). Thus, by identifying these obstacle-free areas, the robot will automatically get hints about where obstacles, but as well objects of interest, might be located. Although we have to restrict our robot to working environments where the floor does neither have bright reflections nor shadows nor texture (big patterns), this method is very reliable, fast to compute (on single 2-D images) and rather conservative, preferring false positives to false negatives.

Robust segmentation is provided by a multilevel image processing algorithm that self-initializes to the floor color in front of the robot and adapts during operation, so that changes in brightness can be compensated to some extent. Built upon this sensor skill of segmenting the image and yielding a polygon of the contour, other skills have been established that enable the robot to detect and recognize, or at least to derive hypotheses about the presence of, objects in the scene relevant for navigation, e.g., junctions, doors and docking stations (e.g., tables shown in Figure 5, right). Upon object recognition, reference points and lines are identified (based on procedural and object knowledge) and subsequently used for tracking.

Tactile sensing. To enable a robot to perceive objects by a sense of touch, in general, tactile sensors are required. Although we are working towards the development of highly integrated and high-resolution tactile sensors to be placed on the gripper surfaces and around the wheelbase, we have found other means of detecting touch events that are helpful in guiding human-robot or robot-environment interaction. By intelligently processing joint angle encoder values and motor currents we are able to provide kinesthetic sensing. Kinesthesis is a "sense mediated by end organs located in muscles, tendons, and joints and stimulated by bodily movements and tensions" [Babcock 1976]. Transferring kinesthetic sensing to the robot for detecting touch events means to detect tensions on the robot structure or torques at the joints that do not result from internal motion requests but most probably from external circumstances.

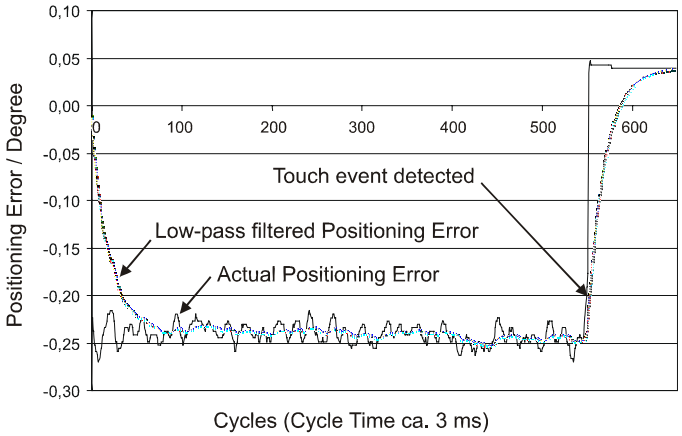

Fig. 6. Positioning error (commanded position- actual position) of the wrist pitch joint during the grasped object's downward movement. As soon as the object touches the table, the positioning error increases, leading to the detection of a touch event

Two kinesthetic sensing skills have been developed: one, for detecting touch events or vibrations that occur on any part of the robot structure; two, for detecting unusual external forces during pre-defined robot motions that are, however, unknown to the sensing skill. While the first skill is being used to interact with people in order to shake hands and to hand over or take objects, the second skill allows to gently place grasped objects onto other objects. In both cases angle encoder values are sampled at a rate of 1 kHz and low-pass filtered to yield a prediction for the next cycle. If a new angle value deviates significantly from the predicted one, a touch event is signaled to the software module that has requested to detect this touch event (Figures 6 and 7).

4.3 Sensorimotor Skills

Combining a few of the above-mentioned motor and sensor skills yields sensor-guided robot motion that leads to goal-directed robot behavior.

Fixating. Visually fixating an environmental point of interest requires, first, a sensor skill that continuously delivers the image coordinates of this point with respect to a predefined fixation point in the image, and second, a motor skill that computes motion control words for the camera head motors in order to minimize the difference between reference and fixation point. A simple proportional control law is used to derive the velocities of the camera head motors: The difference in the y-coordinate between reference and fixation point is used to control the velocity of the tilt axis (elevation angle of the camera) and the difference in the x-coordinate is used to compute the required velocity for the pan-axis.

Wandering around with obstacle avoidance. Wandering around is a basic behavior that can be used by a mobile robot to navigate, explore and map unknown working environments. Our implementation requires the above- mentioned segmentation and fixating skills as well as basic motor skills for controlling the wheelbase. We further assume that the cameras cannot be rotated around their optical axes, i.e., the top of the world is represented in the top of the images, and all objects to be detected rest on the ground plane. After having segmented the image in obstacle-free and occupied areas, the robot can pan its camera head towards those obstacle-free regions that appear to be largest. If such a region is classified as large enough, the steering angle λ of both wheels is set to half of the camera head's pan angle and the robot moves forward while continuously detecting, tracking and fixating this area. The robot will continue its smooth wandering locomotion until it reaches a dead end. Then, the robot stops and invokes a search pattern that scans the floor around the robot. At least in the robot's back (where it has come from) an obstacle-free path should be found. In this case the camera points backwards (pan angle $= 180°$) and the steering angles for both driven wheels are set to $\lambda = 90°$, according to the given control law ($\lambda = /2$). Thus, while the robot tries to move forward, it turns on the spot, and, automatically leaves the dead end situation.

Docking (approaching objects). The main task of a mobile service robot with manipulator arms is to manipulate various objects at different locations. Prerequisite for manipulating objects is to bring them into the working range of arms and grippers, i.e., to navigate the robot sufficiently close to the object to be manipulated. We propose a visual servoing method enabling the robot to approach objects that are in its field of view and to stop in

Fig. 7. Motion sequences for demonstrating tactile sensing skills; top row: interacting with a human and filling a glass with water (from left to right: robot standing by, taking over bottle from a human, taking over glass, visually supervising filling action, handing over glass to a human, handing over bottle); bottom row: taking over various objects from a human (glass, bottle, glue stick, video cassette) and placing them gently onto a tray

front of them at a pose (position and orientation) that is suitable for subsequent manipulation. It is based on the continuous tracking of a predefined reference line that corresponds to a physical edge of the docking object (e.g., a table's front edge) and the fixation of a reference point of the object (e.g., a table corner).

Prerequisites for showing this docking behavior are the segmentation skill (as a basis for object detection), the fixation skill (for keeping the object's reference point at the fixation point in the center of the image), motor skills (for wheelbase control) and cognitive skills (for deriving control words depending on the robot's perceived situation).

The main idea of the docking behavior is to derive the steering angle λ at each moment from the values of the pan angle , the slope of the reference line $m = tan$ in the image and the tilt angle in order to maneuver the robot into its predefined final docking pose (e.g., $= 0°$, $= 0°$, $= {}_{end}$). More details of the docking behavior are beyond the scope of this paper and may be found in [Bischoff, Graefe 1998].

Taking, giving and placing objects. Interaction with people and objects requires tactile sensor skills. In combination with motor skills, such as gross arm positioning, objects can be received from or given to people, or placed onto other objects. Since the robot is not yet skilled enough to visually perceive the current pose of a human hand in order to conform to it, it brings its arm into a configuration where the human user could easily hand over objects or receive them. A touch event will signal that a human has closed the kinematic chain and is willing to receive or give an object. The robot then closes or opens its gripper, respectively.

To place objects onto other objects, the arm with the grasped object has to be grossly positioned first. Since the perceptual abilities are still limited and do not allow to visually guide the manipulator tip with the grasped object to the required location, the arm is fully stretched out first, and, then, commanded to move the first elbow joint down and the wrist joint up with the same velocities, to yield a downward movement of the gripper and, at the same time, to keep it aligned with an assumed horizontal surface (e.g., a table). Supervising all arm modules for a touch event will indicate when either the robot arm or the grasped object have touched something. Figure 6 shows an example of the positioning error of the wrist joint during the arm's downward (and the wrist joint's upward) movement. As soon as a touch event is detected, the arm is halted and the gripper is opened, thus relieving immediately the minimal structural stress upon the robot and the object that occurs when the kinematic chain closes (Figure 7).

4.4 Communicative Skills

The communicative skills of the robots are mostly based on natural language. Natural language may be used by a human to instruct the robot, and by the robot to generate easy-to-understand messages for the user. Commands may be input via voice (via a wireless microphone) or keyboard (directly via a wireless keyboard or indirectly via e-mail messages that may contain multiple commands). The robot displays its messages either on a screen, sends e-mails or generates speech from text.

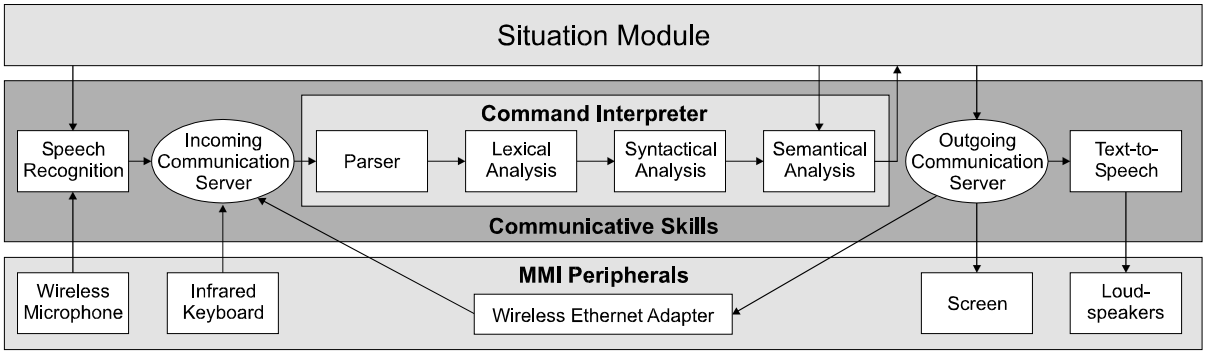

Fig. 8. Visualization of the data flow between the peripherals of the man-machine interface, the communicative skills and the situation module. A communication server handles all incoming and outgoing messages. Incoming messages are interpreted by the command interpreter and subsequently handled by the situation module depending on the robot's actual situation. In turn, the robot's current situation directly influences the speech recognition and the semantical analysis to enhance recognition and interpretation. Outgoing messages are routed to the users via another communication server and appropriate communication channels (e.g., voice, text and graphics display or the Internet)

To enable natural language processing with limited computational resources, a simple grammar has been designed. It is able to cover all the commands, statements and questions that might be given to the robot. Command sentences have a simple structure allowing them to be classified after the first word, thus facilitating the interpretation of the following words. Examples for command sentences are object and action-oriented instructions such as "Go to the kitchen!", or "Grasp the small ball!". Directive instructions such as "Turn around!" or more complex commands like "Turn left at the next intersection!" are supported as well. Intervening commands that do not contain a command verb are partly supported, e.g., "faster" (instead of "move faster"). In this case these adverbs are treated like single command words. Questions are allowed as well, e.g., "What", "Where" and "How". Only a few questions can be answered by the robot so far, e.g., "What is your status?", "Where are you?", "How do I get to the kitchen?" etc. The fixed syntax obviously does not allow an arbitrary reordering of parts of the sentences, e.g., "Take the glass, the big one" or "The glass over there, please take it". Those sentences could however, be admitted, if they had been defined in advance (e.g., using the macro language described below).

Command Interpreter. A command interpreter handles all user input. It consists of a parser, a lexical analysis, a syntactical analysis and a semantical analysis (Figure 8). The parser is fed by a text string that may be provided by the speech recognition module, the e-mail client or directly via a keyboard. It separates the character string into a sequence of words and numbers using space, tabulator and punctuation characters as delimiters. This list is given to the lexical analysis where each word is looked up in a dictionary to obtain its type. Possible types are command verbs (e.g., go, take, place), locations (e.g., office, kitchen, workshop), prepositions (e.g., to, on, onto, in, into), objects (e.g., ball, table, pen) and fill words (e.g., please), just to name a few. Character strings enclosed in quotation marks are treated as one part of a sentence of type "text string". The following syntactical analysis tries to identify the structure of the sentence by comparing the list of types with a list of prototype command sentences that includes all the commands the robot is able to understand. If the comparison is successful the semantical analysis will eventually provide missing words or place holders (such as "it") from the robot's situated knowledge in order to make the command complete.

Speech Recognition. Speech recognition is, in many regards, an unsolved problem. It is especially difficult if speech has to be recognized in the presence of a high level of ambient noise, e.g., inside a moving car or in office environments where the ambient noise includes various machinery sounds, telephones, moving persons or background conversations. Many methods have been proposed to increase the robustness of speech recognition systems, e.g., training in noisy environments or using multiple microphones, e.g., [Matsui et al. 1997].

Since commercial systems able to cope with these problems are not yet available we require, for the time being, the human to use an ordinary wireless microphone to send his commands to the robot. The used speech recognition engine is a commercial product enabling speaker-independent recognition of continuous speech, which means that users may speak to the system naturally, without pauses between words. This is a very important feature because it allows anybody to communicate with the robot without needing any training with the system. The speech recognition engine generates text strings equivalent to the ones that may be entered via the keyboard.

To render the speech recognition more robust, larger word classes such as [object] have been split into several classes, e.g., [object_to_be_manipulated] and [object_used_for_navigation] which are now used as specific arguments of the command words TAKE or GRASP and MOVE or GO. The fewer the number of words per class and the stricter the syntax, the better the results of the speech recognition will be because fewer hypotheses have to be verified. Also, meaningful results are produced even under noisy conditions. Another advantage is that the recognition of a specific grammatical structure can be exploited to detect out-of-vocabulary words. To incorporate these words into the robot's vocabulary a sub-dialogue is initiated that asks the user to spell the unknown word. Spelling is required because words can only be added to the database after converting their notation into phonemes.

Another important way to increase the robustness of the speech recognition system has been the usage of so-called contexts that contain only those grammatical rules and word lists that are needed for a particular situation. Most parts of robot-human dialogues are situated and built around robot-environment or robot-human interactions, a fact which may be exploited to enhance the reliability and speed of the recognition process. When the robot knows what kind of answers it may expect from the user at a given moment it can switch to a situation-specific con text and disable or enable word lists, as appropriate for the current situation. For example, when the robot asks for confirmation, whether it should execute a certain task or not, the answers will be most likely "yes" or "no" and it would make no sense to expect, and to test, other words. By limiting the set of recognizable words or phrases that can actually be expected, the risk of recognition mistakes is reduced considerably.

However, at any stage in the dialogue a few words and sentences not related to the current context must be available to the user, too. These words are needed to "reset" or bootstrap a dialogue, to trigger the robot's emergency stop and to make the robot execute a few other important commands at any time. For example, "Hello, *HERMES*" is used to begin a new dialogue, "Stop" and "Halt" are used for disrupting the robot from its current task, and "Stop listening" and "Continue listening" are used for disabling and enabling the speech recognition engine.

Depending on the prevailing situation and the type of dialogue conducted, various contexts are activated that can be very simple, e.g., allowing only "Yes" or "No" when *HERMES* expects such an answer, or as complex as a navigation context in which multiple phrases and many words exist to allow complex robot control, especially during supervised learning of environmental features. To teach the robot object and place names, a spelling context has been defined that mainly consists of the international spelling alphabet. Since the spelling alphabet has been optimized for ease of use by humans in noisy environments, such as aircraft, it should be well suited for robotic applications, too.

4.5 Data Processing Skills

Data processing skills organize and access the system's knowledge bases. Three types of knowledge bases are being used: an attributed topological map for storing the static characteristics of the environment (for details see [Graefe, Bischoff 1997]), an object data base and a list of missions to accomplish. Depending on his preferences and on the abilities of the robot, the user may define the robot's mission in more or less detail. A mission description may either consist of a detailed list of actions (e.g., elementary behaviors) that are to be executed sequentially by the robot, or only of a single command if the user has enough confidence in the robot's planning abilities. For instance, a route is planned by a data processing skill based on Dijkstra's shortest path algorithm in terms of vision- guided navigation behaviors, e.g., leave home base, turn right, stop at the second door to the right (no coordinates are used).

5 Experiments and Results

We conducted a number of real-world experiments with the humanoid robot *HERMES* to evaluate the concepts presented in the preceding sections. An example that may serve to show the potential of the concepts, but also the limitations of their current implementations, is depicted in Figure 9. The corresponding dialogue with a human user is reprinted in the sequel together with some explanatory remarks:

First, a mission description containing only a single command line is being sent to *HERMES*' e-mail address (HERMES@unibw-muenchen.de). In this e-mail, *HERMES* is instructed to go to the kitchen and to wait for further instructions. To acknowledge the reception of this valid user request, *HERMES* sends back an e-mail message stating that it will execute the commanded mission. One could say that its situation has changed the moment the message arrived (by setting a new transient goal). The required paths from its current location to the kitchen are planned in terms of elementary navigation behaviors by the corresponding data processing skill.

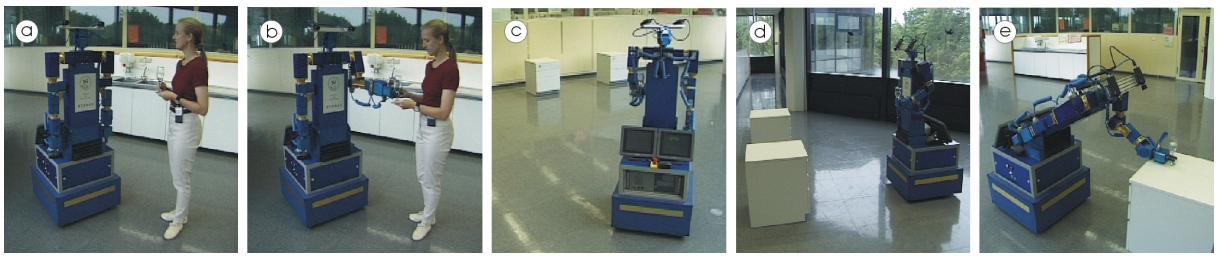

Fig. 9. *HERMES* performing a simple service task upon user request: a) greetings and receiving first instructions (via a wireless microphone); b) receiving a glass from the user using kinesthetic sensing; not shown: further conversation with the user as to what to do next (see text); c) and d) approaching left table using a visual servoing method; e) placing glass onto the table

After having arrived in the kitchen *HERMES* waits for further instructions. When a human arrives *HERMES* detects this event via its visual motion detection skill, turns its head into the direction of the moving person and starts a conversation (also, the human could have addressed *HERMES* in the first place by saying "Hello, *HERMES*!" or "Wake up!", if *HERMES* would not have noticed the person's arrival).

***HERMES*:** "Hello! What can I do for you?"

Human: "Take over glass!"

Simple action- and object-oriented instructions consisting of one command verb and an assigned object are mostly well recognized by the speech recognition system, i.e., they yield a high confidence level. Therefore, the robot will not ask the user for confirmation before actually executing the command. A confirmation is only required if either the confidence level is below a predefined threshold or the instruction would not make any sense in the current context or invoke a number of more complex behaviors that would keep the robot busy for a certain (estimated) amount of time. This mechanism helps to keep the dialog between the robot and the user as fluent as possible.

In this case, the robot checks its data base to figure out what a glass looks like and how it has to be grasped to keep it in an upright and safe pose. Since the robot is not skilled enough to perceive the current pose of the glass (as held by the human) it brings its arm into a configuration where the human can easily hand over the glass. While the gripper is opening the robot says:

***HERMES*:** "Hand over glass, please!"

Now *HERMES* waits for a touch event that it can recognize with its proprioceptive sensors, specifically an involuntary movement of one or more of its joints that most probably results from someone touching its fingers(see section 4.2). After *HERMES* has successfully grasped the glass, it again waits for further instructions.

***HERMES*:** "What else can I do for you?"

Human: "Place it onto the table!"

This sentence invokes a number of different skills. First the robot has to find out what object the user has in mind when referring to "it", then it has to check whether it knows the location of the table and finally it has to approach it. If all parameters are known, the robot breaks down the instruction into elementary skills and asks for confirmation by repeating what it has understood.

***HERMES*:** "I will search for the table, then I will approach the table, then I will place the glass onto the table. Do you want me to do these tasks?"

Human: "Yes, please!"

Since the shape of the table is known, but its actual position is unknown, the robot starts searching for it (based on the segmentation of the image into obstacle-free and obstacle-occupied regions and contour polygon matching, see section 4.2). In this experiment two tables have been placed in the robot's field of view, so that no intensive search would be necessary and the sample dialogue could continue.

***HERMES*:** "I have found two tables. Shall I approach the left or the right table?"

If something unexpected happens or the robot's situation is ambiguous, it informs the user about its current status and then asks for further instructions. The human has to clarify the situation according to the choices given by the robot.

Human: "The left table!"

The speech recognition system only tries to identify the key words "left" or "right" (apart from key phrases to start over or end the current conversation). This improves considerably speech recognition and allows other phrases such as "The left one!" or just the word "left". The final docking to the table is performed using the visual servoing method described in section 4.3. The behavior of placing an object onto another object is again performed utilizing the kinesthetic sense described above.

***HERMES*:** "What else can I do for you?"

Human: "Nothing. Thank you!"

To end the conversation with the robot the user specifies that "nothing" is left that the robot could do for her. Now it is the robot's turn to decide what to do next: waiting for the next mission to accomplish or to obey internal "needs", e.g., to go to the maintenance workshop or battery charger.

This dialogue is a representative example of some of the capabilities of the robot. Natural language recognition is currently restricted to spoken language (besides English also German is possible) with a fixed grammar (mostly imperative sentences that have a relatively simple structure). Nevertheless, the robot shows already fairly cooperative and communicative behaviors as appropriate in its actual situation.

Although presently *HERMES*' visual sensing skills are still based on algorithms that only work under quite restrictive assumptions, they are well suited for studying various control algorithms and validating the proposed situation-oriented behavior-based system architecture. In combination with some basic motor skills more powerful sensorimotor skills can be created. Together with communicative and data processing skills they will lead to goal- directed behavior.

Vision-guided docking yields good results. The robot may start its approach from arbitrary poses and it always stops sufficiently close in front of the docking station and parallel to it. It is important to notice that both the robot's trajectory and the final docking pose are directly derived from sensor data (image features and encoder readings). They are not calculated from distance measurements, kinematic models and inverse perspective transforms with respect to a fixed reference frame (world coordinate system). The method is generic and suitable for all kinds of docking or goal objects where a reference line and a specific point can be defined. Moreover, it may be implemented in a calibration-free or self-calibrating way.

Kinesthetic sensing is sufficiently accurate to provide the robot with a sense of touch. Arbitrary objects can be placed onto other objects without using visual feedback. It remains to be validated to what extent this sense can be used for minimizing damaging effects caused by collisions between *HERMES*' manipulators and other objects. Nevertheless, high resolution tactile sensors for both grippers and around the wheelbase are being developed to enhance the robot's perceptual abilities.

Many more experiments have been carried out and other behaviors have been created based on the elementary skills presented here, such as filling a glass with water (including recognition of the glass and the water level) and following persons. Many pre-programmed control sequences and taught motion patterns can be modified and commanded via a wireless keyboard or a wireless microphone, thus allowing *HERMES* to be teleoperated in all its degrees of freedom for entertainment purposes.

6 Summary

The first part of the paper has addressed some open questions regarding the design and construction of humanoid robots, their need for combined locomotion and manipulation, situated perception und human-friendly communication, advanced safety concepts, and adaptability and learning. A flexible software architecture has been promoted to be able to integrate all before-mentioned components under a coherent framework.

The second part of the paper has been devoted to *HERMES* – a complex robot designed according to an anthropomorphic model. It already integrates various sensor modalities including vision, touch and hearing and displays intelligence and cooperativeness in its behavior and communicates in a user-friendly way, which was demonstrated in experiments conducted with non-expert users. A special kind of behavior-based system architecture has been proposed to control the robot. Its main idea is to select and coordinate the behaviors based on an assessment of the situation being perceived by both the human operator and the robot at a particular moment. This concept places high demands on the robot's sensing and information processing, as it requires the robot to perceive situations and to assess them in real time. A network of microcontrollers, digital signal processors and a single PC, in combination with the concept of skills for organizing and distributing the execution of behaviors efficiently among the processors, is able to meet these demands.

Due to the innate characteristics of the situation-oriented behavior-based approach the robot is able to cooperate with a human and to accept orders that would be given to a human in a similar way. Human-robot communication is based on speech that is recognized speaker-independently without any prior training of the speaker. A high degree of robustness is obtained due to the concept of situation-dependent invocations of grammar rules and word lists called "contexts". Human-robot interaction is facilitated by kinesthetic sensing that consists of intelligently processing angle encoder values and motor currents and enables the robot to hand over and take over objects from a human as well as to gently place objects onto tables or other objects.

7 Conclusions and Outlook

Still much research is needed to endow future personal or service robots with skills enabling their deployment in massive numbers in environments cohabited by humans. Since users of such robots will not be robotic experts, the robot design has to be the more human-like and its control has to be the more human-friendly, the closer the contact with humans will be. Up to date, in many, if not all respects, a human is more versatile, intelligent and adaptable than current embodiments of human-like intelligence. However, a human is still the best model for building personal robots with 'plug-and-play' functionalities, i.e., artificially intelligent humanoids.

It is a very challenging task to bring together expertise in many diverse disciplines, such as electrical and mechanical engineering, computer engineering, and psychology in order to create a robot that closely resembles a human not only in size and shape, but also in sensory and motor skills. Although we are still very far from creating human-like skills and intelligence in an embodied form, methods developed for humanoids could well enhance current service robots and lead to the development of personal robots in the future. In contrast to today's specialized service robots these personal robots could well be used in many different environments (domestic, public and industrial) for a variety of tasks (e.g., elderly care, helping handicapped people, assistance in factories).

As far as *HERMES* is concerned, the development will now focus on integrating various other skills, most importantly skills that enable the robot to manipulate objects guided by vision. The robustness of the methods will be further enhanced by employing calibration-free methods, and learning and adaptation will be used to a greater extent . Since safety is a major concern for all robots intended to work in peopled environments, tactile sensors based on conductive foams have already been developed and will be mounted around the robot's wheelbase, grippers and arms in the near future.

References

Amtec (1997). MoRSE – Modular Positioning and Handling System, Description and Technical Specifications, April 1997.

Arkin, R. C. (1998). Behavior-Based Robotics. MIT Press, Cambridge, MA, 1998.

Atkeson, C. G.; Hale, J.; Kawato, M.; Kotosaka, S.; Pollick, F.; Riley, M.; Schaal, S.; Shibata, T.; Tevatia, G.; Ude, A.; Vjiayakumar, S. (2000): Using Humanoid Robots to Study Human Behavior. IEEE Intelligent Systems magazine (In Press, http://www.erato.atr.co.jp/DB/).

Babcock, P. (1976): Webster's Third New International Dictionary of the English Language, G. & C. Merriam Company, Springfield, MA, 1976.

Bergener, T.; Bruckhoff, C.; Dahm, P.; Janßen, H.; Joublin, F.; Menzner, R. (1997). Arnold: An Anthropomorphic Autonomous Robot for Human Environments. In: H.-M. Groß (Hrsg.): Fortschrittsberichte VDI, Reihe 8, Nr. 663, Workshop SOAVE'97, Ilmenau, September 1997, pp 25-34.

Bischoff, R. (1997). *HERMES* – A Humanoid Mobile Manipulator for Service Tasks. Proc. of the Int. Conference on Field and Service Robotics. Canberra, Australia, December 1997, pp. 508-515.

Bischoff, R. (1999). Advances in the Development of the Humanoid Service Robot *HERMES*. Second International Conference on Field and Service Robotics. Pittsburgh, PA, August 1999, pp. 156-161.

Bischoff, R.; Graefe, V. (1998). Machine Vision for Intelligent Robots. IAPR Workshop on Machine Vision Applications. Makuhari/Tokyo, November 1998, pp. 167-176.

Bischoff, R.; Graefe, V. (1999). Integrating Vision, Touch and Natural Language in the Control of a Situation-Oriented Behavior-Based Humanoid Robot. IEEE Conference on Systems, Man, and Cybernetics, October 1999, pp. II-999 - II-1004.

Bischoff, R.; Graefe, V.; Wershofen, K. P. (1996). Combining Object-Oriented Vision and Behavior-Based Robot Control. Proc. of the Int. Conf. on Robotics, Vision and Parallel Processing for Ind. Automation. Ipoh, Malaysia, pp 222-227.

Brooks, R. A.; Stein, L. A. (1994). Building Brains for Bodies. Autonomous Robots 1(1), pp. 7-25.

Cameron, J. M.; MacKenzie, D. C.; Ward, K. R.; Arkin, R. C.; Book, W. J. (1993). Reactive Control for Mobile Manipula tion. Proceedings IEEE International Conference on Robotics and Automation. Atlanta, GA, May 1993, Vol. 3, pp 784-791.

Dario, P.; Guglielmelli, E.; Laschi, C.; Guadagnini, C.; Pasquarelli, G.; Morana, G. (1995). MOVAID: a new European joint project in the field of Rehabilitation Robotics. http:// www.alfea.it/movaid/Public_Domain_Area/ Papers/Paper1.html, Arts Lab- Scuola Superiore Sant'Anna, Italy.

DLR (2000). DLR Lightweight Robot. Institute of Robotics and Mechatronics. http://www.robotic.dlr.de/LBR/eng_index.html.

Graefe, V. (1989). Dynamic Vision Systems for Autonomous Mobile Robots. Proc. IEEE/RSJ International Workshop on Intelligent Robots and Systems, IROS '89. Tsukuba, pp. 12-23.

Graefe, V. (1995). Object- and Behavior-oriented Stereo Vision for Robust and Adaptive Robot Control. Int. Symp. on Microsystems, Intelligent Materials, and Robots, Sendai, pp. 560-563.

Graefe, V.; Bischoff, R. (1997). A Human Interface for an Intelligent Mobile Robot. 6th IEEE International Workshop on Robot and Human Communication. Sendai, Japan, Sept. 1997, pp 194-197.

Graefe, V.; Maryniak, A. (1998). The Sensor-Control Jacobian as a Basis for Controlling Calibration-Free Robots. IEEE International Symposium on Industrial Electronics, ISIE '98. Pretoria, July 1998, pp. 420-425.

Hanebeck, U.; Fischer, C.; Schmidt, G. (1997). ROMAN: A Mobile Robotic Assistant for Indoor Service Applications. Proc. of IEEE/RSJ Intern. Conference on Intelligent Robots and Systems, IROS '97, September 1997, pp 518-525.

Hirai, K.; Hirose, M.; Haikawa, Y.; Takenaka, T. (1998). The Development of the Honda Humanoid Robot. Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'98). Leuven, Belgium, April 1998.

Honda (1999). The HONDA HUMANOID ROBOT. http://www.honda.co.jp/english/technology/robot/index.html.

Horswill, I. (1994). Visual Collision Avoidance. IEEE/RSJ/GI International Conference on Intelligent Robots and Systems IROS'94, Munich, Germany, September 1994, pp. 902-909.

Hoshino, Y.; Inaba, M.; Inoue, H. (1998). Model and Processing of Whole-body Tactile Sensor Suit for Human-Robot Contact Interaction. Proceedings of 1998 IEEE International Conference on Robotics and Automation, Leuven, Belgium, May 1998, pp. 2281-2286.

Kawamura, K.; Wilkes, D. M.; Pack, T.; Bishay, M.; Barile, J. (1996). Humanoids: Future Robots for Home and Factory. Proceedings of the First International Symposium on Humanoid Robots, Waseda University, Tokyo, Japan, , October 1996, pp. 53-62.

Khatib, O.; Yokoi, K.; Chang, K.; Ruspini D.; Holmberg, R.; Casal A.; Baader A. (1995). Force Strategies for Cooperative Tasks in Multiple Mobile Manipulation Systems. Intern. Symposium of Robotics Research. Munich, October 1995.

Knoll, A.; Hildebrandt, B.; Zhang, J. (1996). Instructing Cooperating Assembly Robots through Situated Dialogs in Natural Language. Research Report 96-01. Universität Bielefeld, Technische Fakultät, Abt. Informationstechnik.

Konno, A.; Nagashima, K.; Furukawa, R.; Nishiwaki, K.; Noda, T.; Inaba, M.; Inoue, H. (1997). Development of the Humanoid Robot Saika. Proc. of IEEE/RSJ Intern. Conference on Intelligent Robots and Systems, IROS '97, September 1997, pp 805-810.

Laengle, T.; Lueth, T. C.; Herzog, G.; Stopp, E.; Kamstrup, G. (1995). KANTRA - a natural language interface for intelligent robots. In U. Rembold and R. Dillmann, eds., Intelligent Autonomous Systems IAS-4, pp. 365--372. IOS Press, 1995.

Lueth, T. C.; Nassal, U. M., Rembold, U. (1995). Reliability and Integrated Capabilities of Locomotion and Manipulation for Autonomous Robot Assembly. Journal on Robotics and Autonomous Systems, 14 (1995), pp 185-198.

Matsui, T.; Asoh, H.; Asano, F. (1997). Map Learning of an Office Conversant Mobile Robot, Jijo-2, by Dialogue-Guided Navigation. Proceedings of the International Conference on Field and Service Robotics. Canberra, Australia, December 1997, pp. 230-235.

Milde, J.-T.; Peters, K.; Strippgen, S. (1997). Situated Communication with Robots. First International Workshop on Human-Computer Conversation. Bellagio, Italy, July 1997.

Morita, T.; Iwata, H.; Sugano, S. (1999). Design of Body Mechanism for Realizing Human-Robot Symbiosis. Second International Symposium on Humanoid Robots. Tokyo, Japan, October 1999, pp. 121-128.

Nagatani, K.; Yuta, S. (1998). Autonomous Mobile Robot Navigation Including Door Opening Behaviour - System Integration of Mobile Manipulator to Adapt Real Environment. In A. Zelinsky (ed.): Field and Service Robotics. Springer, London 1998, pp. 195-202.

Pauly, M.; Finke, M.; Peters, L.; Beck, K. (1999). Control and Service Structure of a Robot Team. IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS'99, Kyongju, Korea, October 1999, pp. 1069-1074.

Schlegel, C.; Wörz, R. (1999). The Software Framework SmartSoft for Implementing Sensorimotor Systems. IEEE/RSJ International Conference on Intelligent Robots and Systems, Kyongju, Korea, October 1999, pp. 1610-1616.

Schraft, R. D.; Schmierer, G. (1998). Serviceroboter – Produkte Szenarien, Visionen. Springer-Verlag, Berlin, 1998 (in German).

Thórisson, K. R. (1999). A Mind Model for Multimodal Communicative Creatures and Humanoids. International Journal of Applied Artificial Intelligence, 13(4-5), 449-486.

Thrun, S.; Fox, D.; Burgard, W. (1998). A Probabilistic Approach to Concurrent Mapping and Localization for Mobile Robots. Machine Learning 31, pp. 29-53 and Autonomous Robots 5, pp. 253-271, (joint issue).

Torrance, M. C. (1994). Natural Communication with Robots. Master Thesis, MIT, Department of Electrical Engineering and Computer Science, Cambridge, MA, 1994.

Ward, K.; Zelinsky, A. (1998). An Exploratory Robot Controller which Adapts to Unknown Environments and Damaged Sensors. In A. Zelinsky (ed.): Field and Service Robotics. Springer, London 1998, pp. 456-463.

Yamamoto, Y. (1994). Control and Coordination of Locomotion and Manipulation of a Wheeled Mobile Manipulator. Dissertation, University of Pennsylvania, August 1994.

Zhang, J.; Collani, Y. v.; Knoll, A. (1998). Development of a Robot Agent for Interactive Assembly. Proceedings of 4thInternational Symposium on Distributed Autonomous Robotic Systems, Karlsruhe, Germany, 1998.