# Past, Present and Future of Intelligent Robots

Volker Graefe and Rainer Bischoff
Intelligent Robots Lab, LRT 6
Bundeswehr University Muenchen
85577 Neubiberg, Germany
http://www.UniBw-Muenchen.de/campus/LRT6

## Abstract

Some fundamental characteristics of past, present and future robots are reviewed. In particular, the humanoid robot *HERMES*, an experimental robotic assistant of anthropomorphic size and shape, and the key technologies developed for it, are introduced. *HERMES* interacts dependably with people and their common living environment. It understands spoken natural language (English, French and German) speaker-independently, and can, therefore, be commanded by untrained humans.

*HERMES* can see, hear, speak, and feel, as well as move about, localize itself, build maps of its environment and manipulate various objects. In its dialogues and other interactions with humans it appears intelligent, cooperative and friendly. In a long-term test (6 months) at a museum it chatted with visitors in natural language in German, English and French, answered questions and performed services as requested by them.

## 1 Introduction

Machines that resemble humans or animals have fascinated mankind for thousands of years, but only in the 16th century technology and craftsmanship became sufficiently advanced both in Europe and in Japan to allow the construction of automated dolls. What we call robots today are machines that incorporate at least some computational intelligence, and such machines have existed only for a few decades.

The most wide-spread robots today are industrial robots. They are useful and important for the production of goods, but they are not very intelligent. With the advent of more powerful computers more intelligent artificial creatures could be realized, including some autonomous vehicles and service robots.

In the future we will see "personal robots" that will entertain, comfort and serve people in their private lives and homes. While presently robotic servants or butlers exist only in the form of early prototypes in a few research laboratories, they are expected to become as ubiquitous as PCs in the future.

There is no precise definition, but by general agreement a robot is a programmable machine that imitates the actions or appearance of an intelligent creature, usually a human. To qualify as a robot, a machine has to be able to do two things: one, get information from its surroundings, and two, do something physical, such as move or manipulate objects. Robots can be huge and massive 50 meters long machines or little tiny manipulators in micro- or nanometer space. They can be intelligent and autonomously (unpredictably) act on their environment, or dumb machines repeatedly making the same predictable and precise motions without a pause, or something in-between. They are propelled by wheels or tracks, move snake-like or have legs; they work in laboratories, offices or museums, act in outer space or swim in the deep sea. Robots are made to accomplish dirty, dull or dangerous work, and more recently, to entertain and to be played with. They construct, assemble, cut, glue, solder, weld, paint, inspect, measure, dig, demine, harvest, clean, mow, play soccer and act in movies. This “multi-cultural society” has grown in recent years to more than one million “inhabitants”.

### 1.1 Ancient Robots

Probably the oldest mentioning of autonomous mobile robots may be found in Homer’s Iliad (written circa 800 B.C.). According to this source, Hephaistos, the Greek god of smiths, fire and metalworking, built 20 three-legged creatures (tripods) “with golden wheels beneath the base of each that of themselves they might enter the gathering of the gods at his wish and again return to his house” (book 18, verse 375). They are described as being powerful and intelligent, with ears and voices, willing to help and work for him [Homer 800 B.C.]. – Details regarding their technology are left to the imagination of the reader.

Mechanical animals that could be animated by water, air and steam pressure were constructed by Hero of Alexandria in the first century B.C. [Woodcroft 1851]. Much later, depending on dexterous manufacturing knowledge for clockworks starting in the 16th century, skilled craftsmen in Western Europe succeeded to design anthropomorphic devices that could imitate a human’s movements or behaviors in general. Mechanical dolls performed simple life-like acts, such as drawing, writing short phrases or playing music [Heyl 1964].

Japanese craftsmen of the 18th century created many varieties of automated mechanical dolls, *karakuri*, that could perform such acts as drawing an arrow from a quiver, shoot it from a bow, and display pride over the good shot. Another famous *karakuri* could bring a tea cup to a guest over distances of about 2 m (size of a *tatami* mat). When the guest removed the cup from the tray, the doll ceased to move forward, turned around and returned to its starting place [Nipponia 2000]. What makes those *karakuri* particularly fascinating is that their mechanisms are usually constructed entirely from wood.

Modern *karakuri* combine a beautiful and artistic appearance with sophisticated computer-controlled mechanics inside. Figure 1 shows as an example a *karakuri* created by the artist Yuriko Mudo and on display in a department store in Nagoya station. Such dolls may nowadays be seen in many public places, hotel lobbies and restaurants in Japan.

## 1.2 Industrial Robots

Other successors to the ancient robots are today's industrial robots. While they may be more useful, they are certainly less artistic. More than one million industrial robots are working in the factories of the world, producing many of those goods which we like to consume or use every day. While these robots are an important source of our prosperity, they have no intelligence and very little sensory abilities. They can operate only in carefully prepared environments and under the supervision of experts. For safety reasons they must stop moving whenever a safety barrier is violated by a person or an object, even if the robot is not nearby.

## 1.3 Autonomous Mobile Robots

In the 1960s and 1970s some ambitious researchers at Stanford University, Jet Propulsion Laboratory and Carnegie Mellon University created a novel kind of robots: computer-controlled vehicles that ran autonomously in their laboratories and even outside with a video camera as the main sensor [Nilsson 1969], [Moravec 1980]. Due to the limited computing power and insufficient vision technology of the time, the speed of those early vehicles was only about 1 m in 10-15 min, and the environment had to be carefully prepared to facilitate image interpretation.

In 1987 technology had advanced to the point that an autonomous road vehicle could follow a road at a speed of 96 km/h, a world record at that time [Dickmanns, Graefe 1988]. In 1992 the objects that are relevant for road traffic situations could be recognized in real time from within a moving vehicle [Graefe 1992], making it possible for an autonomous driverless vehicle to mix with ordinary vehicles in ordinary freeway traffic. Although most major automobile companies now operate autonomous cars in their research laboratories, decades will pass before such vehicles will be sold to the public.

In recent years another kind of robots has appeared in the market. Unlike industrial robots, their purpose is not the production of goods in factories, but the delivery of various services, so far mainly in the areas of floor cleaning [Endres et al. 1998], mail delivery [Tschichold 2001], lawn-mowing [Friendly Robotics 2003], giving tours in a museum [Nourbakhsh et al. 1999], [Thrun et al. 2000] and surgical assistance [Integrated Surgical Systems 2001]. They have been employed in environments where they may, or even have to, come into contact with the public, and some of them actually interact with people. They can, to a very limited extent, perceive their environment and they display traces of intelligence, e.g., in navigation and obstacle avoidance. Combined with their slow speed of motion this allows some of them to operate safely in the vicinity of ordinary humans. All these service robots, as they are called, have the following characteristics in common (a few exceptions exist):

**Figure 1:** Modern computer-controlled *karakuri* “Ciélo arpéggío” with four dolls. The doll on the right plays an instrument as the other ones dance to the tune. (From [Mudo 2003])

- Each one of them is a specialist, able to deliver only one kind of service in only one kind of environment.
- Their sensory and cognitive abilities and their dependability are barely sufficient for accomplishing their given task most of the time.
- They are of a more or less experimental nature and have not yet proven their cost effectiveness.

Much R&D effort is being spent to overcome these deficiencies and it is hoped that service robots will eventually be economically as important as industrial robots are today.

## 1.4 Personal Robots

A novel kind of robots is currently evolving. While industrial robots produce goods in factories, and service robots support, or substitute, humans in their work places, those novel “personal robots” are intended to serve, or accompany, people in their private lives and share their homes with them. Two types of personal robots have so far emerged: One type comprises robots that are intended to make people feel happy, comfortable or less lonely or, more generally speaking, to affect them emotionally; these robots usually cannot, and need not, do anything that is useful in a practical sense. They may be considered artificial pets or – in the future – even companions. Therefore, they are also called personal robotic pets or companions. The most famous one is AIBO, sold in large numbers by Sony since 1999. Weighing about 2 kg it resembles in its appearance and some of its behaviors a miniature dog. The other type of personal robot is intended to do useful work in and around peoples' homes and eventually evolve into something like artificial maids or butlers. Such robots may be called personal robotic servants or assistants.

In many developed societies the fraction of elderly people is growing and this trend will continue for at least several decades. Consequently, it will be more and more difficult to find enough younger people to provide needed services to the elderly ones, to help them with their households, to nurse them and even to just give them company. We may hope that personal robots will help to alleviate these problems. Looking at it from a different point of view, and also considering the fact, that many of those elderly people are fairly wealthy and have relatively few heirs for whom they might want to save their wealth, personal robots promise to create large and profitable markets for technology-oriented companies. It is not surprising that major companies, such as Fujitsu, NEC, Omron, Sanyo, Sony and Honda are developing and marketing personal robots [Fujitsu 2003],[NEC 2001], [Omron 2001], [Sanyo 2002], [Fujita & Kitano 1998], [Sakagami et al. 2002].

**Figure 2:** Humanoid experimental robot *HERMES*; mass: 250 kg; size: 1.85 m · 0.7 m · 0.7 m

Technologically, pet robots are much less demanding than servant robots. Among the reasons are that no hard specification exists for what a pet robot must be able to do, and that many deficiencies that a cute pet robot might have may make it even more lovable in the eyes of its owner. Assisting a pet robot in overcoming its deficiencies may actually be an emotionally satisfying activity. A servant robot, on the other hand, simply has to function perfectly all the time. Even worse: while a maid will be forgiven her occasional mistakes if she offers sincere apologies, no technology is available for implanting the necessary capacities for sincerity, feeling of guilt and compassion in a robot. In fact, marketable servant robots are far beyond our present technology in many respects and all personal robots that have been marketed are pet robots.

Pet robots have already demonstrated their indirect usefulness in systematic studies. For instance, Shibata and coworkers [Wada et al. 2003] have carried out rehabilitation experiments in various hospitals with a white furry robot seal called Paro (the name comes from the Japanese pronunciation of the first letters of ‘personal robot’). Paro has 7 degrees of freedom, tactile sensors on the whiskers and most of its body, posture and light sensors, and two microphones. It generates behaviors based on stimulation (frequency, type, etc.), the time of day and internal moods. Paro has one significant advantage over artificial cats and dogs: people usually do not have pre-conceived notions about seal behavior and are unfamiliar with their appearance, and thus people easily report that the interaction with Paro seems completely natural and appropriate. The seal’s therapeutic effect has been observed in hospitals and among elderly. During several interaction trials in hospitals carried out over several months, researchers found a marked drop in stress levels among the patients and nurses. Nurses of an elderly day care center reported that the robot both motivated elderly people and promoted social communication.

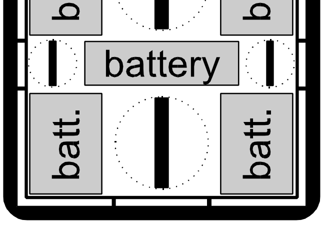

**Figure 3:** *HERMES’* omnidirectional undercarriage with active (large) and passive (small) wheels, bumpers and batteries

Servant robots, on the other hand, exist only in the form of early prototypes in a few research laboratories, and then often not even as complete robots. In some cases only a head, or the image of a simulated head on a screen, exists, in other cases only a torso with a head and arms, but without the ability of locomotion.

In the remainder of this paper we will introduce one of these prototypes, the humanoid experimental robot *HERMES* that we have developed to advance the technology of servant robots (Figure 2). What makes it special is the great variety of its abilities and skills, and the fact that its remarkable dependability has actually been demonstrated in a long-term test in a museum where it interacted with visitors several hours a day for six months.

## 2 The Humanoid Robot *HERMES*

### 2.1 Overview

With its omnidirectional undercarriage, body, head, eyes and two arms *HERMES* has 22 degrees of freedom and resembles a human in height and shape. Its main exteroceptive sensor modality is monochrome vision.

In designing it we placed great emphasis on modularity and extensibility of both hardware and software [Bischoff 1997]. It is built from 25 drive modules with identical electrical and similar mechanical interfaces. Each module contains a motor, a Harmonic Drive gear, a microcontroller, power electronics, a communication interface and some sensors. The modules are connected to each other and to the main computer by a single bus. The modular approach has led to an extensible design that can easily be modified and maintained.

Both camera “eyes” may be actively and independently controlled in pan and tilt degrees of freedom. Proprioceptive sensors add to *HERMES’* perceptual abilities. A multimodal human-friendly communication interface built upon natural language and the basic senses – vision, touch and hearing – enables even non-experts to intuitively interact with, and control, the robot.

### 2.2 Hardware

*HERMES* has an omnidirectional undercarriage with 4 wheels, arranged on the centers of the sides of its base (Figure 3). 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°) (Figure 4). The work space extends up to 120 cm in front of the robot. Each arm is equipped with a two-finger gripper that is sufficient for basic manipulation experiments.

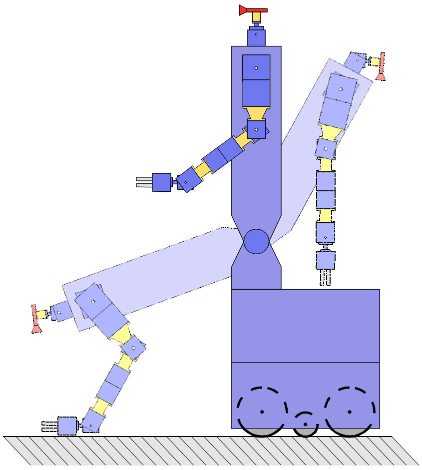

**Figure 4:** A bendable body greatly enlarges the work space and allows the cameras to be always in a favorable position for observing the hands.

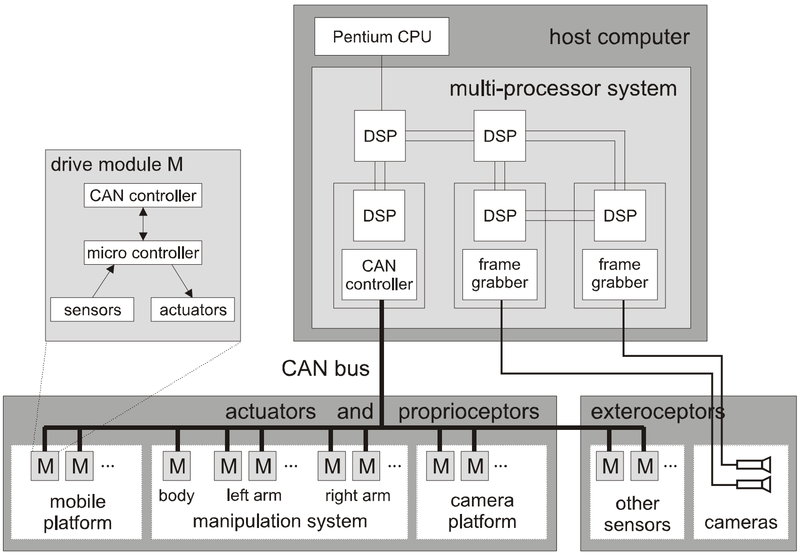

**Figure 5:** Modular and adaptable hardware architecture for information processing and robot control

Main sensors are two video cameras mounted on independent pan/tilt drive units ("eye modules"), in addition to the pan/tilt unit ("neck module") that controls the common "head" platform. The cameras can be moved with accelerations and velocities comparable to those of the human eye.

A hierarchical multi-processor system is used for information processing and robot control (Figure 5). 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 ruggedized, but otherwise standard industrial PC. Sensor data processing (including vision), situation recognition, behavior selection and high-level motion control are performed by the DSPs, while the PC provides data storage, Internet connection and the human interface.

A robot operating system was 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.

## 3 System and Software Architecture

### 3.1 Overview

Overall control is realized as a finite state automaton that does not allow unsafe system states. It is 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, received via e-mail, entered via keyboard, or result from a spoken dialogue. 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 or spoken, respectively, in natural language and passed to a parser and an interpreter. If a command cannot be understood, is under-specified or ambiguous, the situation module tries to complement missing information from its situated knowledge or asks the user via its communicative skills to provide it.

Several of the fundamental concepts developed earlier by our laboratory were implemented in *HERMES* and contribute to its remarkable dependability and versatility, e.g., an object-oriented vision system with the ability to detect and track multiple objects in real time [Graefe 1989] and a calibration-free stereo vision system [Graefe 1995]. The sensitivities of the cameras can be individually controlled for each object or image feature. Several forms of learning let the robot adapt to changing system parameters and allow it to start working in new environments immediately. Moreover, speaker-independent speech recognition for several languages and robust dialogues, at times augmented by appropriate gestures, form the basis for various kinds of human-robot interaction [Bischoff, Graefe 2002].

### 3.2 System Architecture

Seamless integration of many – partly redundant – degrees of freedom, numerous behaviors 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 communication 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.

Figure 6 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 (man-machine, 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.

### 3.3 Skills

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*** control 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, such as undercarriage, arms, body and head, leads to a simple interface structure and allows an easy generation of pre-programmed motion patterns. Motor skills are mostly implemented at the microcontroller 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*** encapsulate the access to one or more sensors and provide the situation module with proprioceptive or exteroceptive data. Sensor skills are implemented on those DSPs that have direct access to digitized sensor data, especially digitized images.

***Sensorimotor skills*** combine both sensor and motor skills to yield sensor-guided robot motions, e.g., vision-guided or tactile and force-and-torque-guided robot motions.

***Communicative skills*** pre-process user input and generate a valuable feedback for the user according to the current situation and the given application scenario.

***Data processing skills*** are responsible for organizing and accessing the system's knowledge bases. 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 of our system architecture which bases upon the concepts of situation, behavior and skill see [Bischoff, Graefe 1999].

***Cognitive skills*** are realized by the situation module in the form of situation assessment and behavior selection, based on data and information fusion from all system components. Moreover, the situation module provides general system management and is responsible for planning appropriate behavior sequences for reaching given goals, i.e., it coordinates and initializes 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.

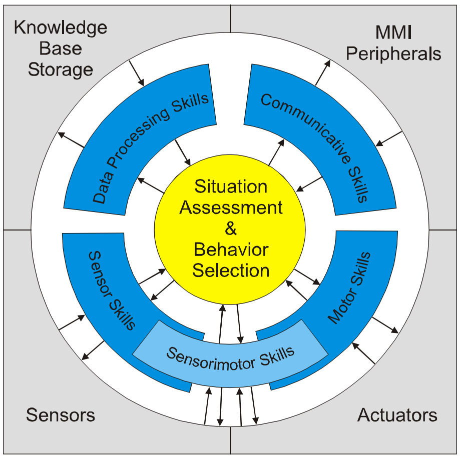

**Figure 6:** *HERMES'* system architecture, based on the concepts of situation, behavior and skill

## 4 Communication and Learning

### 4.1 Overview

It is a basic ability of any personal robotic servant to interact and communicate with humans. Usually the human partners of a servant robot will wish to use its services, but they are not necessarily knowledgeable, or even interested, in robotics. Also, they will not be motivated to modify their habits or their homes for the benefit of a robotic servant. Therefore, the robot must communicate in ways that humans find natural and intuitive, and it must be able to learn the characteristics of its users and its environment. For reasons of cost no expert help will be available when these characteristics change, or when the robot is to begin to work in a new environment. Communication and learning abilities are, therefore, crucial for a servant robot.

### 4.2 Communication

**Speaker-independent voice recognition.** *HERMES* understands natural continuous speech independently of the speaker, and can, therefore, be commanded in principle by any non-dumb human. This is a very important feature, not only because it allows anybody to communicate with the robot without needing any training with the system, but more importantly, because the robot may be stopped by anybody via voice in case of emergency. Speaker-independence is achieved by providing grammar files and vocabulary lists that contain only those words and provide only those command structures that can actually be understood by the robot. In the current implementation *HERMES* understands about 60 different command structures and 350 words, most of them in each of the available three languages English, French and German.

**Robust dialogues for dependable interaction.** Most parts of robot-human dialogues are situated and built around robot-environment or robot-human interactions, a fact that has been exploited to enhance the reliability and speed of the recognition process by using so-called contexts. They contain only those grammatical rules and word lists that are needed for a particular situation. However, at any stage in the dialogue a number of words and sentences not related to the current context are 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.

Obviously, there are some limitations in our current implementation. One limitation is that not all utterances are allowed, or can be understood, at any moment. The concept of contexts with limited grammar and vocabulary does not allow for a multitude of different utterances for the same topic. In general, speech recognition is not sufficiently advanced, and compromises have to be accepted in order to enhance the recognition in noisy environments. Furthermore, in our implementation it is currently not possible to track a speaker's face, gestures or posture. This would definitely increase the versatility and robustness of human-robot communication.

### 4.3 Learning

**Learning by doing.** Two forms of learning are currently being investigated. They both help the robot to learn by actually doing a useful task: One, to let the robot automatically acquire or improve skills, e.g., grasping of objects, without quantitatively correct models of its manipulation or visual system (*autonomous learning*). Two, to have the robot generate, or extend, an attributed topological map of the environment over time in cooperation with human teachers (*cooperative learning*).

The general idea to solve the first learning problem is simple. While the robot watches its end effector with its cameras, like a playing infant watches his hands with his eyes, it sends more or less arbitrary control commands to its motors. By observing the resulting changes in the camera images it "learns" the relationships between such changes in the images and the control commands that caused them. After having executed a number of test motions the robot is able to move its end effector to any position and orientation in the images that is physically reachable. If, in addition to the end effector, an object is visible in the images, the end effector can be brought to the object in both images and, thus, in the real world.

Based on this concept a robot can localize and grasp objects without any knowledge of its kinematics or its camera parameters. In contrast to other approaches with similar goals, but based on neural nets, no training is needed before the manipulation is started [Graefe 1999].

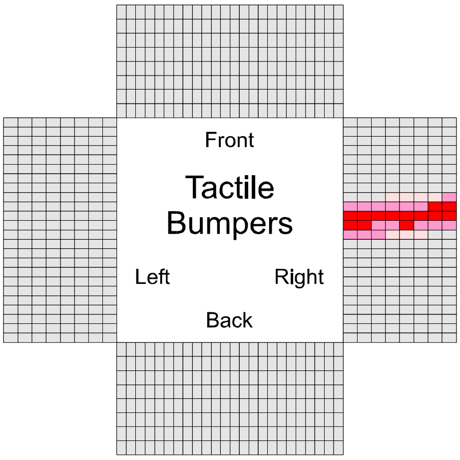

**Figure 7:** Sensor image of tactile bumpers after touching the corner of two adjacent walls while the robot was trying to turn around it; color coding: light grey value = no touch, the darker the color the higher the exerted forces during touch; the sensor image outer row to inner row correspond to a covered area from 40 - 320 mm above the ground on the undercarriage.

The general idea to solve the second learning problem is to let the robot behave like a new worker in an office with the ability to explore, e.g., a network of corridors, and to ask people for reference names of specific points of interest, or to let people explain how to get to those points of interest. The geometric information is provided by the robot's odometry, and relevant location names are provided by the persons who want the robot to know a place under a specific name. In this way the robot learns quickly how to deliver personal services according to each user's individual desires and preferences, especially: how do (specific) persons call places; what are the most important places and how can one get there; where are objects of personal and general interest located; how should specific objects be grasped? The ability to link, e.g., persons' names to environmental features, requires several databases and links between them in order to obtain the wanted information, e.g., whose office is located where, what objects belong to specific persons and where to find them.

Many types of dialogues exist to *cooperatively* teach the robot new knowledge and to build a common reference frame for subsequent execution of service tasks. For instance, the robot's lexical and syntactical knowledge bases can easily be extended, firstly, by directly editing them (since they are text files), and secondly, by a dialogue between the robot and a person, that allows to add new words and macro commands during run-time.

To teach the robot names of persons, objects and places that are not yet in the database (and, thus, cannot be understood by the speech recognition system), a spelling context has been defined that mainly consists of the international spelling alphabet. This alphabet has been optimized for ease of use by humans in noisy environments, such as aircraft, and has proved its effectiveness for our applications as well, although its usage is not as intuitive and natural as individual spelling alphabets or as a more powerful speech recognition engine would be.

## 5 Experiments and Results

Since its first public appearance at the Hannover Fair in 1998 where *HERMES* could merely run (but still won "the first service robots' race"!) quite a number of experiments have been carried out that prove the suitability of the proposed methods. Of course, we performed many tests during the development of the various skills and behaviors of the robot and often presented it to visitors in our laboratory. The public presentations made us aware of the fact that the robot needs a large variety of functions and characteristics to be able to cope with the different environmental conditions and to be accepted by the general public.

In all our presentations we experienced that the robot's anthropomorphic shape encourages people to interact with it in a natural way. One of the most promising results of our experiments is that our calibration-free approach seems to pay off, because we experienced drifting of system parameters due to temperature changes or simply wear of parts or aging. These drifts could have produced severe problems, e.g., during object manipulation, had the employed methods relied on exact kinematic modeling and calibration. Since our navigation and manipulation algorithms only rely on qualitatively (not quantitatively) correct information and adapt to parameter changes automatically, the performance of *HERMES* is not affected by such drifts.

Tactile sensing also greatly improves the system's dependability. Figure 7 shows an example of the tactile bumper sensors' response in case of an accident. In this simple contact situation *HERMES* tries to continue to deliver its service, e.g., to transport an object, and does not wait until a human has solved the problem. In such a simple case the robot would drive backwards, modify the steering angle and try again. More complex contact situations (2 or more contact locations) still require, for safety reasons, the help of a human.

**Cooperative Learning Experiment**

*In this experiment HERMES and a person are standing in the room commonly referred to as "kitchen". HERMES has been recently introduced to this and some other rooms and only knows a few people so far. It is in a state of waiting for a new instruction within the (default) service context.*

**Person: "Hello!"**

*HERMES has to find out who is requesting a service since it is not yet able to recognize people or faces.*

***HERMES*: "Hello. My name is *HERMES*. What is your name, please?"**

**Person: "My name is Rainer."**

*HERMES cannot find the person's name in its database, and therefore, switches to the spelling context.*

***HERMES*: "I am sorry, but I do not know you. Please spell your name!"**

**Person: "Romeo Alpha India November Echo Romeo."**

*HERMES switches to the confirmation context.*

***HERMES*: "Your name is Rainer. Is this correct?"**

**Person: "Yes!"**

*HERMES extends its database of persons and switches back to the service context.*

***HERMES*: "Hello Rainer. Pleased to meet you. What can I do for you?"**

**Person: "Take this glass. Bring it into my office."**

*These sentences invoke a number of different skills. First HERMES takes over the glass the person currently holds in his hand. Then, the robot resolves the pronouns "it" and "my" from its situated knowledge. However, the robot does not yet know where the person's office is located. HERMES has to ask the user to provide the missing information.*

***HERMES*: "Please describe the way to your office or give me the room number."**

*HERMES switches to the navigation context.*

**Person: "The room number is 2455!"**

*HERMES switches back to the service context, and starts searching for the room. Alternatively, the person could have given instructions on how to reach the room, such as "Leave the kitchen through the door in front of you. My office is located at the second door to the left!". After having actually found the room, HERMES extends its database of known locations, and marks the room as one of Rainer's offices.*

**Figure 8:** Excerpt from a dialogue between a human and *HERMES* to transport an object to another room. In its course, *HERMES* learns more about its environment and stores this knowledge in several databases for later reference (e.g., the attributed topological map shown in Figure 9). It should be noted how often contexts are switched, depending on the robot's expectations. This improves the speech recognition considerably.

The dialogue depicted in Figure 8 may serve as an example how robots and people in general could build a common reference frame in terms preferred by the user in their shared working environment. Whenever a command is incomplete (missing command arguments) or ambiguous (too many arguments or imprecise description), a specific dialogue is initiated to resolve the problem. It is important to note that it is always the robot (except in an emergency) who is in charge of the current dialogue and the flow of information towards the user.

Autonomously or through dialogues with people, the robot is able to build an attributed topological map of its environment (Figure 9). Since *HERMES* is using only vision for its navigation it is limited by its relatively poor perception (when compared to humans). Nevertheless, the situation-oriented and skill-based system architecture, in addition to the camera's active sensitivity control, enables a navigation performance that is more than adequate for our office building environment. Combined visual and tactile sensing is only in its early stages. We expect the robot to perform even more dependably when these senses are fully integrated and combined.

In the sequel we concentrate on demonstrations that we performed outside the familiar laboratory environment, namely in television studios, at trade fairs and in a museum where *HERMES* was operated by non-experts for an extended period of time. Such demonstrations, e.g., in television studios, subject the robot to various kinds of stress. First of all, it might be exposed to rough handling during transportation, but even then it should still function on the set. Second, the pressure of time during recording in a TV studio requires the robot to be dependable; program adaptation or bug-fixing at the location is not possible.

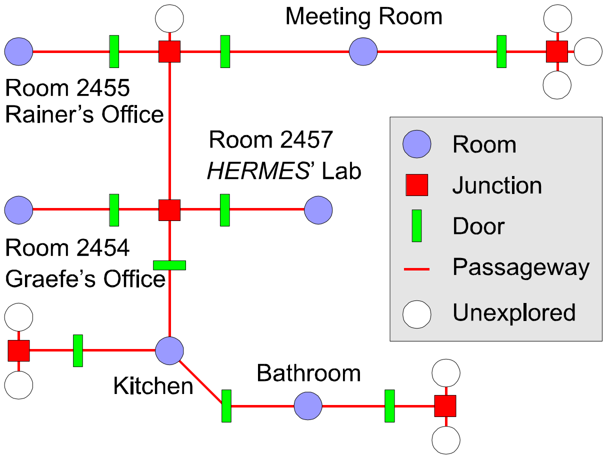

**Figure 9:** Attributed topological map built by the robot by autonomous exploration or with help of human teachers through dialogues (e.g., the dialogue depicted in Figure 8). The robot learns how persons call (specific) places and how the places are connected via passageways. Multiple names are allowed for individual locations, depending on users' preferences. Geometric information does not have to be accurate as long as the topological structure of the network of passageways is preserved. (The map has been simplified for demonstration purposes. It deviates significantly in terms of complexity, but not in general structure, from the actual map being used for navigation around the laboratory.)

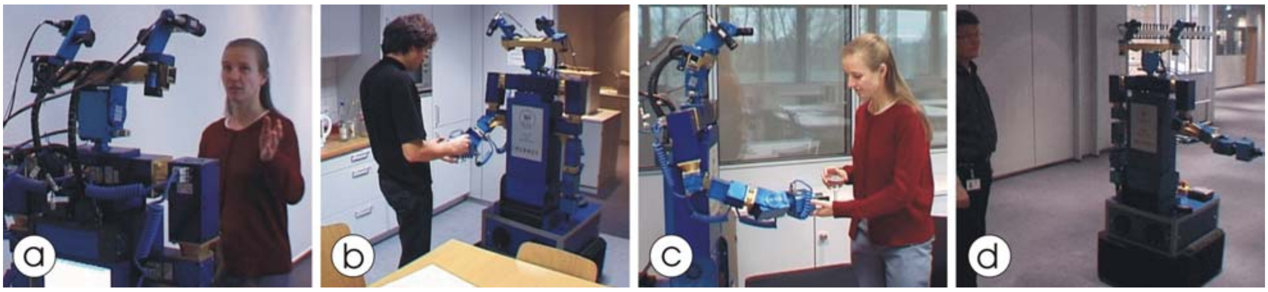

**Figure 10:** *HERMES* executing service tasks in the office environment of the Heinz Nixdorf MuseumsForum: (a) dialogue with an *a priori* unknown person with *HERMES* accepting the command to get a glass of water and to carry it to the person's office; (b) asking a person in the kitchen to hand over a glass of water; (c) taking the water to the person's office and handing it over; (d) showing someone the way to a person's office by combining speech with gestures (head and arm) generated automatically.

*HERMES* performed in TV studios a number of times and we have learned much through these events. We found, for instance, that the humanoid shape and behavior of the robot raise expectations that go beyond its actual capabilities, e.g., the robot is not yet able to act upon a director's command like a real actor (although sometimes expected!). It is through such experiences that scientists get aware of what "ordinary" people expect from robots and how far, sometimes, these expectations are missed.

Trade fairs, such as the Hannover Fair, the world's largest industrial fair, pose their challenges, too: hundreds of moving machines and thousands of people in the same hall make an incredible noise. It was an excellent environment for testing the robustness of *HERMES'* speech recognition system.

Last, but not least, *HERMES* was field-tested for more than 6 months (October 2001 - April 2002) in the Heinz Nixdorf MuseumsForum (HNF) in Paderborn, Germany, the world's largest computer museum. In the special exhibition "Computer.Brain" the HNF presented the current state of robotics and artificial intelligence and displayed some of the most interesting robots from international laboratories, including *HERMES*.

We used the opportunity of having *HERMES* in a different environment to carry out experiments involving all of its skills, such as vision-guided navigation and map building in a network of corridors; driving to objects and locations of interest; manipulating objects, exchanging them with humans or placing them on tables; kinesthetic and tactile sensing; and detecting, recognizing, tracking and fixating objects while actively controlling the sensitivities of the cameras according to the ever-changing lighting conditions.

*HERMES* was able to chart the office area of the museum from scratch upon request and delivered services to *a priori* unknown persons (Figure 10). In a guided tour through the exhibition *HERMES* was taught the locations and names of certain exhibits and some explanations relating to them. Subsequently, *HERMES* was able to give tours and explain exhibits to the visitors. *HERMES* chatted with employees and international visitors in three languages (English, French and German). Topics covered in the conversations were the various characteristics of the robot (name, height, weight, age, ...), exhibits of the museum, and actual information retrieved from the World Wide Web, such as the weather report for a requested city, or current stock values and major national indices. *HERMES* even entertained people by waving a flag that had been handed over by a visitor; filling a glass with water from a bottle, driving to a table and placing the glass onto it; playing the visitors' favorite songs and telling jokes that were also retrieved from the Web (Figure 11).

## 6 Conclusions and Outlook

By integrating various sensor modalities, including vision, touch and hearing, a robot may be built that displays intelligence and cooperativeness in its behavior and communicates in a user-friendly way. This was demonstrated in experiments with a complex robot designed according to an anthropomorphic model.

The robot is basically constructed from readily available motor modules with standardized and viable mechanical and electrical interfaces. Due to its modular structure, *HERMES* is easy to maintain, which is essential for

**Figure 11:** *HERMES* performing at the special exhibition "Computer.Brain", instructed by commands given in natural language: taking over a bottle and a glass from a person (not shown), filling the glass with water from the bottle (a); driving to, and placing the filled glass onto, a table (b); interacting with visitors (here: waving with both arms, visitors wave back!) (c)

system dependability. A simple but powerful skill-based system architecture is the basis for software dependability. It integrates visual, tactile and auditory sensing and various motor skills without relying on quantitatively exact models or accurate calibration. Actively controlling the sensitivities of the cameras makes the robot's vision system robust with respect to varying lighting conditions (albeit not as robust as the human vision system). Consequently, safe navigation and manipulation, even under uncontrolled and sometimes difficult lighting conditions, were realized. A touch-sensitive skin currently covers only the undercarriage, but is in principle applicable to most parts of the robot's surface.

*HERMES* understands spoken natural language speaker-independently, and can, therefore, be commanded by untrained humans. This concept places high demands on *HERMES'* sensing and information processing, as it requires the robot to perceive situations and to assess them in real time. A network of microcontrollers and digital signal processors embedded in 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, *HERMES* 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". A kinesthetic sense, based on intelligently processing angle encoder values and motor currents greatly facilitates human-robot interaction. It enables the robot to hand over, and take over, objects from a human as well as to smoothly place objects onto tables or other objects.

*HERMES* interacts dependably with people and their common living environment. It has shown robust and safe behavior with novice users, e.g., at trade fairs, television studios, in our institute environment, and in a long-term experiment carried out at an exhibition and in a museum's office area.

In summary, *HERMES* can see, hear, speak, and feel, as well as move about, localize itself, build maps and manipulate various objects. In its dialogues and other interactions with humans it appears intelligent, cooperative and friendly. In a long-term test (6 months) at a museum it chatted with visitors in natural language in German, English and French, answered questions and performed services as requested by them.

Although *HERMES* is not as competent as the robots we know from science fiction movies, the combination of all before-mentioned characteristics makes it rather unique among today's *real* robots. As noted in the introduction, today's robots are mostly strong with respect to a single functionality, e.g., navigation or manipulation. The results achieved with *HERMES* illustrate that many functions can be integrated within one single robot through a unifying situation-oriented behavior-based system architecture. Moreover, they suggest that testing a robot in various environmental settings, both short- and long-term, with non-experts having different needs and different intellectual, cultural and social backgrounds, is enormously beneficial for learning the lessons that will eventually enable us to build dependable personal robots.

## References

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

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

**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. (2002):** Dependable Multimodal Communication and Interaction with Robotic Assistants. Proceedings 11th IEEE International Workshop on Robot and Human Interactive Communication (ROMAN 2002). Berlin, pp 300-305.

**Dickmanns, E.D.; Graefe, V. (1988):** Dynamic Monocular Machine Vision; and: Applications of Dynamic Monocular Machine Vision. Machine Vision and Applications 1 (1988), pp 223-261.

**Endres, H.; Feiten, W.; Lawitzky, G. (1998):** Field test of a navigation system: Autonomous cleaning in supermarkets. Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 1998), Vol. 2, pp. 1779-1784.

**Friendly Robotics (2003):** Robomower. Owner Operating and Safety Manual. http://www.friendlyrobotics.com, last visited on March 22, 2003.

**Fujita, M.; Kitano, H. (1998):** Development of an autonomous quadruped robot for robot entertainment. Journal of Autonomous Robots, Vol. 5, No. 1, pp 7-18.

**Fujitsu (2003):** Fujitsu, PFU Launch Initial Sales of MARON-1 Internet-Enabled Home Robot to Solutions Providers in Japan Market. Press Release, Fujitsu Limited and PFU Limited, Tokyo, March 13, 2003, http://pr.fujitsu.com/en/news/2003/03/13.html, last visited on May 31, 2003.

**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. (1992):** Visual Recognition of Traffic Situations by a Robot Car Driver. Proceedings, 25th ISATA; Conference on Mechatronics. Florence, pp 439-446. (Also: IEEE International Conference on Intelligent Control and Instrumentation. Singapore, pp 4-9.)

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

**Graefe, V. (1999):** Calibration-Free Robots. Proceedings, the 9th Intelligent System Symposium. Japan Society of Mechanical Engineers. Fukui, pp. 27-35.

**Heyl, E. G. (1964):** Androids. In F. W. Kuethe (ed.): The Magic Cauldron No. 13, October 1964. Supplement: An Unhurried View of AUTOMATA. Downloaded from http://www.uelectric.com/pastimes/automata.htm, last visited on April 22, 2003.

**Homer (800 BC):** The Iliad. In Gregory Crane (ed.): The Perseus Digital Library. Tufts University, Medford, MA 02155, http: //www.perseus.tufts.edu/cgi-bin/ptext?doc= Perseus: text:1999.01.0134: book=1:line=1, last visited on May 5, 2003.

**Integrated Surgical Systems (2001):** Integrated Surgical Systems Announces the Sale of Robodoc® Surgical Assistant System. Press Release, Davis, CA, USA. Available at: http://www.robodoc.com/eng/press_release. html, last visited on May 31, 2003.

**Moravec, H. (1980):** Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover. Doctoral dissertation, Robotics Institute, Carnegie Mellon University, May, 1980.

**Mudo, Y. (2003):** http://www2.neweb.ne.jp/wc/MUDO-ART/, last visited on June 5, 2003.

**NEC (2001):** NEC Develops Friendly, Walkin' Talkin' Personal Robot with Human-like Characteristics and Expressions. Press Release NEC Corporation, Tokyo, March 21, 2001, http://www.nec.co.jp/press/en/0103/2103.html, more information available at: NEC Personal Robot Center, http://www.incx.nec.co.jp/robot/, last visited on May 31, 2003.

**Nilsson, N. (1969):** A Mobile Automaton: An Application of Artificial Intelligence Techniques. Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 1969). Washington D.C., May 1969. Reprinted in: S. Iyengar; A. Elfes (eds.): Autonomous Mobile Robots, Vol. 2, 1991, IEEE Computer Society Press, Los Alamitos, pp. 233-244

**Nipponia (2000):** 21st Century Robots. Will Robots Ever Make Good Friends? In Ishikawa, J. (ed.): Nipponia No. 13, June 15, 2000, Heibonsha Ltd, Tokyo.

**Nourbakhsh, I.; Bobenage, J.; Grange, S.; Lutz, R.; Meyer, R.; Soto, A. (1999):** An Affective Mobile Educator with a Full-time Job. Artificial Intelligence, 114(1-2), pp. 95-124.

**Omron (2001):** "Is this a real cat?" – A robot cat you can bond with like a real pet – NeCoRo is born. News Release, Omron Corporation, October 16, 2001, http://www.necoro.com/newsrelease/index.html, last visited on March 29, 2003.

**Sakagami, Y.; Watanabe, R.; Aoyama, C.; Matsunaga, S.; Higaki, N.; Fujimura, K. (2002):** The Intelligent ASIMO: System Overview and Integration. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, EPFL, Lausanne, Switzerland, October 2002, pp. 2478-2483.

**Sanyo (2002):** tmsuk and SANYO reveals new and improved "Banryu" home-robot. News Release, tmsuk Co., LTD. & SANYO Electric Co., Ltd., Tokyo, November 6, 2002, http://www.sanyo.co.jp/koho/hypertext4-eng/0211news-e/1106-e.html, last visited on May 31, 2003.

**Thrun, S.; Beetz, M.; Bennewitz, M.; Burgard, W.; Cremers, A. B.; Dellaert, F.; Fox, D.; Hähnel, D.; Rosenberg, C.; Roy, N.; Schulte, J.; Schulz, D. (2000):** Probabilistic algorithms and the interactive museum tour-guide robot minerva. Intern. Journal of Robotics Research, Vol. 19, No. 11, pp. 972-999.

**Tschichold, N., Vestli, S., Schweitzer, G. (2001):** The Service Robot MOPS: First Operating Experiences. Robotics and Autonomous Systems 34:165-173, 2001.

**Wada, K; Shibata, T.; Saito, T.; Tanie, K. (2003):** Psychological and Social Effects of Robot Assisted Activity to Elderly People who stay at a Health Service Facility for Aged. Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2003), May 2003, Taipei, Taiwan, to appear.

**Woodcroft, B. (1851):** The Pneumatics of Hero of Alexandria, from the Original Greek Translated for and Edited by Bennet Woodcroft, Professor of Machinery in University College, London; Taylor Walton and Maberly, Upper Gower Street and Ivy Lane Paternoster Row, London, 1851; http://www.history.rochester.edu/steam/hero, last visited on April 22, 2003.