A Bionic Approach to the Design of Dependable Intelligent Robots

Volker Graefe and Rainer Bischoff

Intelligent Robots Laboratory
Bundeswehr University Munich
85577 Neubiberg, Germany

Abstract

Dependability issues are the main obstacle towards introducing 'intelligent' robots to service applications in everyday environments and factories to any major extent. Since the dependability of complex technical systems is largely a consequence of fundamental design decisions we present an approach to robot design that should lead to the construction of robots that are dependable and exhibit intelligent behavior.

A key point of our approach is that intelligent robots, and especially their sensing and information processing, should implement to the greatest practical extent concepts that may be observed in living organisms. After all, nature has created extremely complex, and at the same time extremely dependable, organisms. Following this approach we have designed and constructed a humanoid robot, *HERMES*, that has proved both its dependability and its intelligence in several public short term presentations and in a long-term test.

HERMES' system architecture, several of its skills and the underlying design principles and key technologies are introduced, and some long- and short-term real-world experiments carried out with the robot are presented.

1 Introduction

Industrial robots have already taken over most dangerous and stupid tasks in manufacturing with high accuracy and unbeatable process quality. In the interest of simplicity and cost-effectiveness they do not rely on sensors, but on well-known and prefabricated environments. For future industrial robots it will, however, be indispensable to use a variety of sensors and to 'intelligently' process their signals for sensing their environment to be able to closely and safely collaborate with humans. This collaboration will ultimately lead to a new manufacturing paradigm enabling the efficient production of customizable goods with varying geometries and materials that are only required in small quantities down to lot size 1. Similar requirements exist for personal robots that may be expected to provide in the future a variety of services for humans in their homes. Such robots will combine many sensors and actuators and very complex software. To make them really *helpful* to humans and *productive* in manufacturing requires functionality and dependability levels that are not available today.

In this paper we argue that we need to abstract some good design principles and concepts from nature to achieve higher functionality and dependability levels. The reason is that nature has created highly complex, efficient and dependable systems in the form of organisms since the very beginning of life on earth. Design and function of organisms have been optimized under evolutionary pressure over billions of years, a small step at a time. It would be foolish to not apply nature's solutions to today's engineering problems, since design constraints and objectives are very similar, e.g., functionality, optimization and cost-effectiveness.

This bionic approach has guided us in developing several key technologies that are important for dependable and intelligent robots, such as sensing and perception; localization and navigation; human-robot communication and interaction; adaptability and learning.

2 A Realized Example

To evaluate the potential of the bionic approach for actual applications we have constructed a humanoid robot, *HERMES* (Figure 1) that implements those technologies in one rather complex system. Several of the fundamental concepts developed by our Laboratory have been implemented in *HERMES* and contribute to its remarkable dependability.

In general, all robot components were conceived in such a way that they allow a wide field of possible experiments and the execution of a multitude of tasks. Several bionic design principles [Bischoff, Graefe 2004] have been applied during the requirement analysis and construction phases of our robot; first, to the robot's main hardware components – *body*, *sensors*, *brain*, *output devices* and *energy provision* – and second, to the *system and control architecture* that provides the "glue" to connect and control all components through the robot's *"nervous system"*.

From a hardware perspective, our *robot's body* basically consists of frames, links and joint actuators and can be subdivided into locomotion platform (lower body), manipulation system (upper body) and head. Following our bionic taxonomy of senses the *sensors of our robot* can be classified into extero-, proprio- and interoceptors that provide information about exterior and interior system states. The sensor types should be carefully selected and placed onto or into the body structure in such a way that their data allow in principle the recognition in real time of the (usually complex) situations that the robot may encounter.

Sensors and actuators have to be connected via wires (the robot's "nerves") to the information processing and control system that can be considered as the *robot's brain*. Last but not least, energy is provided through batteries, and transferred via wires (the robot's "veins") to all consumers.

Details regarding the implementation of some of these concepts in a realized technical system will be given in the following sections. Videos showing *HERMES* and its subsystems in operation may be found at [Videos].

2.1. System 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 vision. Both camera "eyes" may be actively and independently controlled in pan and tilt degrees of freedom. The robot is mainly constructed from 25 mechatronic building blocks, or drive modules, with identical mechanical and electrical interfaces. The robot's anthropomorphic body can be subdivided in three main subsystems: undercarriage (lower body), manipulation system (upper body), and (sensor) head.

Lower Body (Undercarriage)

In designing *HERMES* we decided to use wheels and no legs, but we designed an omnidirectional undercarriage that gives the robot a similar freedom of mobility as a human, at least on a flat ground. At first sight using wheels seems to constitute a departure from our bionic approach, but we had two reasons for our decision:

Figure 1: Versatile personal robot assistant *HERMES*; size: 1.85 m x 0.7 m x 0.7 m; mass: 250 kg. The black material covering the undercarriage on all 4 sides is plastic foam that serves as a protective bumper and as part of an imaging tactile sensor.

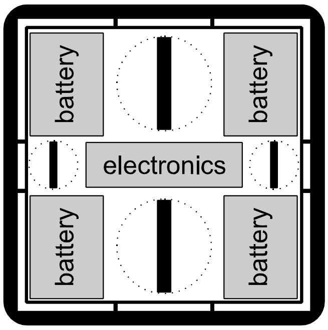

Figure 2: *HERMES'* undercarriage, surrounded by a tactile bumper; driven and steered wheels in the front/back; smaller caster wheels left/right.

1. Legged locomotion is an independent area of research that we do not pursue in our Laboratory.

2. Where it is practical, wheeled locomotion is much more energy-efficient than legged locomotion, which gives the robot a much longer endurance with a single battery charge.

What makes the undercarriage design particularly interesting is that it allows for many different propulsion configurations (Figures 2 and 3): If only the front wheel is steered, the robot behaves like a tricycle or car; if only the rear wheel is steered, it can maneuver like a fork lift; and if both wheels are steered, *HERMES* exhibits omnidirectional capabilities, including diagonal drive modes and rotating on the spot. Furthermore, *HERMES'* locomotion platform can also be configured as differential drive by turning both driven wheels by 90° (to align them in parallel). To use *HERMES* in this case, its upper body would have to be manually lifted and rotated by 90° before mounting it again to have the front side look forward and not sideways.

Conductive foam covering the undercarriage on all four sides serves as a protective bumper and as part of an imaging tactile sensor. Apart from providing locomotion to the robot the undercarriage is also carrying the robot's energy supply in the form of heavy batteries (Figure 2). They power the robot for several hours and make for a low center of gravity and, thus, for a high degree of stability.

Upper body (Manipulation System)

The upper body of *HERMES* consists of a base frame and two articulated arms attached to a torso that can bend forward and backward. Each arm has 6 DOF and is constructed from four rotary actuator modules, a 2-DOF wrist module and a two-finger gripper. They are mechanically connected through cylindrical, conical and angular coupling elements. Compared to humans, *HERMES'* shoulder joints are missing one

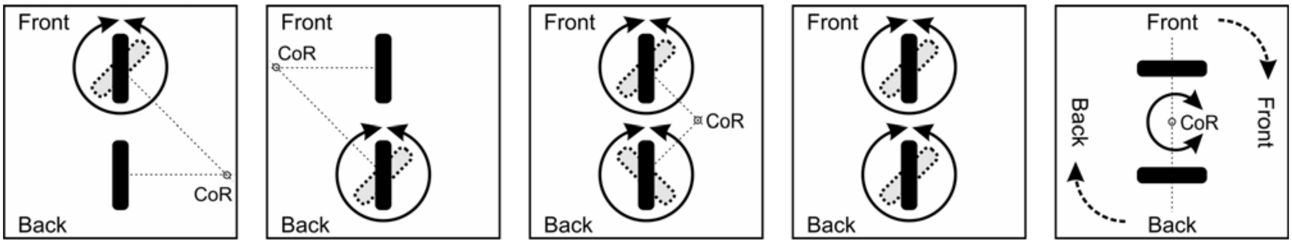

Figure 3: Various drive modes may be realized (car, fork lift, circular, diagonal, differential)

degree of freedom. This may be substituted by an appropriate rotation of the mobile platform, but due to the modular concept an additional DOF could have also been included later, if experiments would have required a redundant manipulator.

The waist module of the bendable upper body is single-axis and situated at a height of 70 cm. Due to its high reduction gear it can only rotate at a maximum speed of 4°/s. Forward and backward bending motions can reach 130° from the upright pose forward and 90° backward, respectively. With the given proportions of the bendable torso and arms a very large workspace from the rear part of the loading area extending up to 1.20 m in front of the robot can be achieved. This allows for manipulating objects in the back of the robot as well as picking up objects from the ground or bending the body over tables to reach out for objects lying far away from the table's edges, which mimics human manipulation abilities (Figure 4).

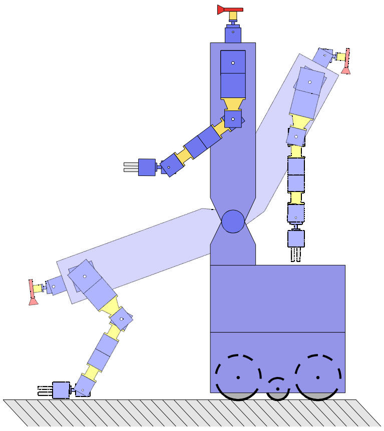

Figure 4: Preliminary design study (1996) to demonstrate the significantly enlarged work space gained by a bendable body. The sensor head remains always in a favorable position for observing end effector activities.

Sensor Head

HERMES' head is carrying the robot's main exteroceptive sensors: two video cameras and a single microphone for capturing sound, especially speech. The head is based on a neck with two degrees of freedom that is identical to the wrist modules of the robot's arms. The neck's kinematics is based on a pan-tilt serial chain, with the first vertical rotation axis (maximum speed 180°/s) about *HERMES*' body flange, and then a pitch motion (maximum speed 90°/s) about the horizontal axis. Although this design brings the stronger tilt actuator on the top of the weaker pan actuator, this arrangement is also found in the human neck. The "Atlas" joint is the first cervical (neck) vertebra, which is just under the head and enables tilting motions. The "Axis" joint is the second cervical vertebra, about which the Atlas rotates. This design has two main advantages: First, a rotation about the pan axis does not change the viewing angle of the world, i.e., the top in the image is still depicting the top of the currently seen part of the world – as opposed to many other pan-tilt systems which tilt first and rotate then. A second advantage – from our point of view – originates from having the pitch motion axis *not* passing through the camera lenses (as with most other binocular active vision camera heads), but is instead about 10 cm below them. This offset lets the cameras translate forward or backward during a tilting motion, enabling *HERMES* to see down on its body and on the ground in front of it, or on the loading area in its back.

Speed and acceleration of the eye modules were designed to be comparable to the human oculomotor system with the main aim of being able to implement saccadic and smooth pursuit eye movements for an active object tracking. The camera eyes can be moved very fast with accelerations up to 10000°/s^2 and velocities up to 800°/s. The common pan-tilt neck unit achieves accelerations of 860°/s^2 and velocities of 215°/s. In contrast to a human, *HERMES* can (independently) pan its neck and eyes more than ±180°, and tilt as much as the installed cameras allow without colliding with the body structure.

Compared to the human mobility range, and also other humanoid robot heads, this is an unparalleled motion range for a binocular head-eye system. It especially allows vision-guided manipulation of objects behind the robot's back, e.g., at the loading area. The independent pan of the eye modules – another important difference to the human and many other robot heads – enables *HERMES* to look sideways with one eye while still being able to look forward with the other. This is useful in narrow passages or corridors to locate, e.g., door plates at the side wall while moving forward along the wall guided by vision.

It should be noted that the anthropomorphic shape of *HERMES'* manipulation and head systems are of vital importance for the implemented sensing and motor skills. As can been from Figure 4 it is advantageous to mount a vision system on a head-like sensor carrier platform above a dual-arm and bendable manipulation system, not only to be able to increase the working range to human-comparable dimensions, but also to have the camera *always* in the right location for a close vision-based supervision of the area in which objects have to be manipulated.

Brain (Information Processing System)

The selection of the right information processing system is crucial to the overall success of a complex robotic system. For reasons of maintainability and extensibility we followed a hierarchical and modular concept choosing dedicated processors on each hierarchical level (Figure 5). A decentralized and hierarchically structured multi-processor system seemed to be the best solution for such a demanding task. Depending on the required computational power, data links, and available peripheral devices different types of processors were chosen for every hierarchical level.

On the highest hierarchical level an operator should be able to enter tasks and supervise the whole system. Normal users and bystanders should also be able to describe more complex service tasks and to give simple commands. A personal computer and its peripherals (infrared keyboard, microphone and speakers, WLAN module and two TFT displays) were selected to implement a human-friendly man-machine interface that is suitable for both experienced and inexperienced users because of the numerous peripheral input and output devices available.

On the next lower hierarchical level commands are generated for sensor-based actions of the robot, i.e., motion commands for the undercarriage, the manipulator system, and the head and sensor system. A homogeneous network of multiple digital signal processors (DSP) deemed to be adequate for this task because its computational power can be flexibly adjusted to the sensor, cognitive and control tasks, and it can be easily upgraded by adding more computational nodes as the complexity of the task or number of degrees of freedom increase.

The lowest hierarchical level is formed by microcontrollers that control actuators and pre-process sensor data. They are mostly embedded in the drive modules and connected via a standardized bus providing real-time communication.

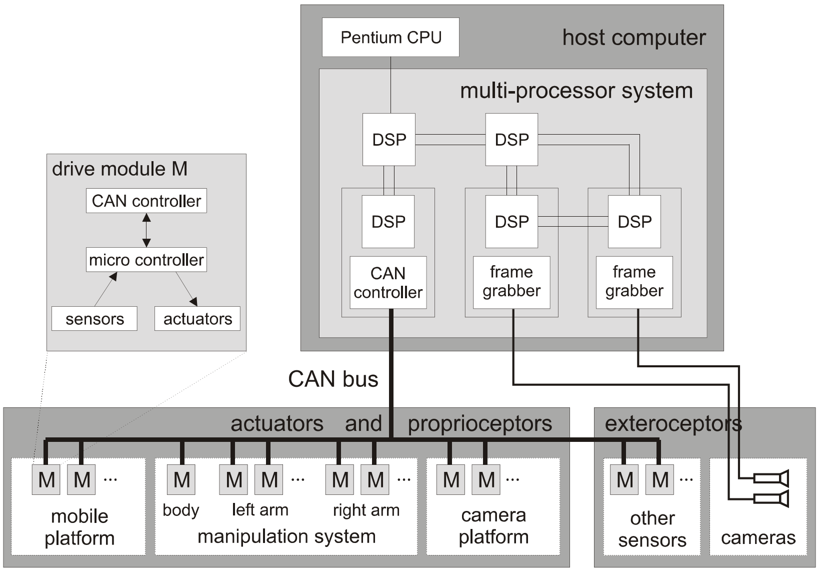

Figure 5: Modular and adaptable hardware architecture for information processing and robot control.

2.2 Sensors

According to the bionic approach sensing should be based exclusively on those senses which have proved their effectiveness in nature. Therefore, vision – the sensor modality that predominates in nature – is also the main sensor modality for our robots. In addition to it, tactile sensing and hearing, also wide-spread in nature, greatly improve *HERMES'* overall sensing abilities.

Vision

The human eye is able to cope with an extremely large range of light intensities. Within a single scene the apparent brightness of visible objects may differ by 5 orders of magnitude (e.g., a sun-lit snow covered road surface vs. a barely visible object in a dark tunnel entrance). Between different scenes (e.g., bright daylight vs. a moonless night) the eye can adapt to a range of light intensities that spans 11 orders of magnitude. Such abilities have evolved because they are advantageous for survival, but no technical image sensors with similar abilities exist. The intra-scene dynamic range of solid-state cameras extends over little more than 2 orders of magnitude at best.

Nevertheless, a sense of vision can be realized in many different ways by means of photo-electronic devices. After weighing the advantages and disadvantages of various technological options we decided to select a CCD imager, analogue video signals and monochrome (black-and-white) cameras. To increase the robustness of the vision system the robot's two CCD cameras have been modified to allow their integration time, and thus their sensitivity, to be controlled by the vision system. Their sensitivities may be adjusted within 20 ms by almost 3 orders of magnitude. This allows the vision system to image in successive video frames bright and dark features in the same scene with good contrast. The robot is then able to observe objects even under unfavorable and changing lighting conditions by maintaining a high contrast around tracked features or keeping an average grey level within a region of interest. Automatic gain control which is usually based on an average grey level within the *entire* image does not yield satisfactory results because it cannot cope with the high differences in brightness of natural scenes which leads to an over- or underexposure of regions of interest or tracked objects.

Tactile Sense

The bodies of humans and most animals are covered by a skin that carries countless sensors for force, temperature, vibration etc. Although it is presently impossible to make an artificial skin of similar complexity and functionality we have covered part of *HERMES'* surface with an artificial skin that gives the robot a kind of tactile sense. This "skin" is based on conductive foam that serves two purposes: one, it damps accidental and unwanted impacts between the robot and humans or environmental objects, and two, it allows to identify the contact locations of, and the forces exerted by, the touched objects. Contact points and forces are sensed via a dense grid of electrodes underneath the foam. Pressing the foam results in a higher conductivity of the material (lower resistance, respectively). The resistance between each electrode and a ground plane is continuously measured (sample rate 50 Hz) and evaluated by dedicated micro controllers.

A touch-sensitive bumper surrounding the robot's undercarriage has already been realized (Figures 1, 2 and 6). Furthermore, two two-finger grippers that are completely covered by this conductive foam were developed and integrated. In the future it is hoped to cover the whole robot structure with this kind of touch-sensitive skin.

Auditory Sense

Until now, the robot's auditory sense is only used to pick up speech. A single microphone attached to the sensor head suffices for first experiments in this direction. At a next stage of the robot's development a microphone array should be fixed to its head to allow the direction from which sound waves arrive to be perceived. Such an ability would be useful to let the robot focus its attention to certain events in the environment and to react to unforeseen events. It may also be a basis for discriminating between several persons who speak simultaneously.

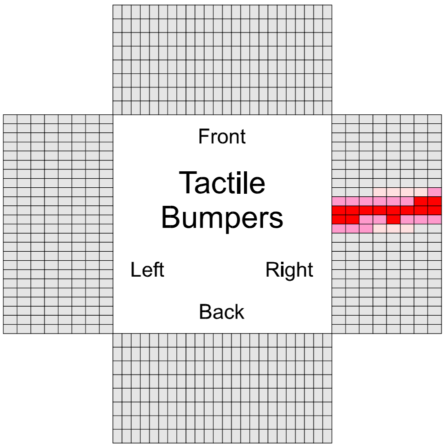

Figure 6: Sensor image of tactile bumpers that cover the four sides of the undercarriage after touching the corner of two adjacent walls while the robot was trying to turn around it; color coding: light grey value = no touch, the darker the color the higher the exerted forces; the sensor image outer row to inner row correspond to a covered area from 40 - 320 mm above the ground around the undercarriage.

Proprioceptive Sense

Proprioceptors are sensors that directly gain information about body and limb movement and posture. All sensors that are used by *HERMES* to gain such bodily information are embedded in its mechatronic building blocks. The most important proprioceptors are very high resolution angle encoders that measure the angles of each of *HERMES'* joints. Their values are fed directly to the joint's micro controller to realize position and velocity control, but their data are also available to higher-level processors in the system. Other sensors that are incorporated in each drive module and used for proprioceptive purposes are motor current sensors. The magnitude of the current applied to reach a certain position or velocity can be used to determine the resistance a motor has to overcome. Hence, current sensing can be regarded as equivalent to force or torque sensing.

So far *HERMES* uses neither inertial sensors (gyroscopes or acceleration sensors) that would correspond to a vestibular sense, nor strain gauges to determine bodily stress. Instead, current and angle measurements within the robot's joints are on the one hand sufficient to compute the posture based on a kinematic model, and an the other hand to derive structural stress information that may be used for kinesthetic sensing (see below).

2.3 Skill-Based System Architecture

HERMES' system architecture was derived from a qualitative model of human information processing and insights gained from psychological literature dealing with skill acquisition, human performance and motor learning. It is based on the concepts of skill and behavior – concepts that are used to describe human and animal behavior as well.

HERMES' system architecture allows the 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. The system architecture has been specified in a way that would allow other developers to integrate their functionality with ease.

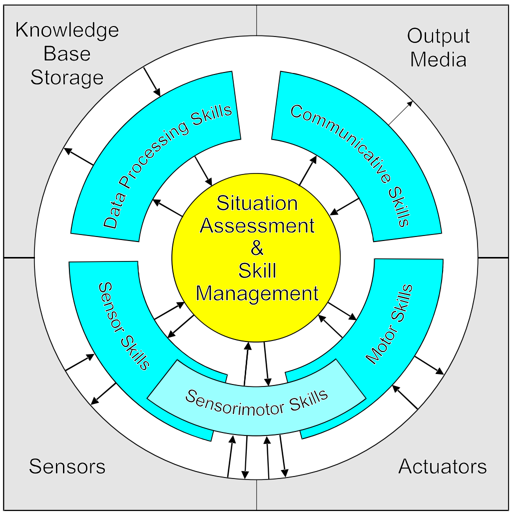

Figure 7: System architecture of an intelligent robot based on the concepts of situation, behavior and skill

Figure 7 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 hardware components – sensors, actuators, knowledge base storage and MMI (man-machine, machine-machine interface) peripherals.

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), and it is responsible for planning appropriate behavior sequences to reach given goals, 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.

Skills 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. Some skills have direct access to the hardware components and, thus, actually realize behavior primitives. For a more profound theoretical discussion of our system architecture which bases upon the concepts of situation, behavior and skill, see [Bischoff, Graefe 1999].

3. Skills

3.1 Classification

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: *Sensor skills* encapsulate the access to one or more sensors and provide the situation module with proprioceptive or exteroceptive data. *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 that form robot subsystems, such as undercarriage, arms, body and head, leads to a simple interface structure, and allows an easy generation of pre-programmed motion patterns. *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 inputs and generate outputs to the users. They are crucial for providing feedback 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.

Videos showing examples of most of the skills mentioned in the sequel may be found at [Videos].

3.2 Calibration-Free Approach

In realizing the sensor, motor and sensorimotor skills we follow a calibration-free approach [Graefe 1999], [Graefe 2008] wherever it is practical. Inspired by nature it deviates significantly from the classical control-theory-based approach, and it gives our robots a high degree of robustness against parameter changes and of adaptability to changing environmental characteristics.

According to the classical approach, robot control is model-based. Numerical models of the kinematics and dynamics of the robot and of the external objects that the robot should interact with, as well as quantitative sensor models, are the basis for controlling the robot's motions. The main advantage of model-based control is that it lends itself to the application of classical control theory and, thus, may be considered a straight-forward approach. The weak point of the approach is that it breaks down when there is no accurate quantitative agreement between reality and the models. Differences between models and reality may come about easily; an error in just one of the many coefficients that are part of the numerical models can suffice. Among the possible causes for discrepancies are initial calibration errors, aging of components, changes of environmental conditions, such as temperature, humidity, electromagnetic fields or illumination, maintenance work and replacement of components, to mention only a few. Consequently, most industrial robots work only in carefully controlled environments and need frequent maintenance (including repetitive calibration), in addition to a cumbersome and expensive initial calibration.

Organisms, on the other hand, are robust and adapt easily to changes of their own conditions and of the environment. They never need any calibration, and they normally do not know the values of any parameters related to the characteristics of their "sensors" or "actuators". Obviously, they do not suffer from the shortcomings of model-based control, which leads us to the assumption that they use something other than quantitative measurements and numerical models for controlling their motions. It seems that their motion control is based on a holistic assessment of situations for the selection of behaviors to be executed. We are convinced that robotics should benefit from following a similar approach.

3.3 Sensor Skills

Vision

One of the most needed sensor skills is to detect and observe objects in the robot's surroundings. Among the objects that a mobile robot needs to observe 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., [Graefe 1990] or [Horswill 1994]). Thus, by identifying these obstacle-free areas, the robot will automatically get hints about where obstacles or other objects of interest might be located. Although we have to restrict our robot to working environments where the floor has neither bright reflections nor shadows nor large texture patterns, this method is very reliable, fast to compute (using only 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. Fast image processing is a prime prerequisite for continuous and smooth robot motion control. Therefore, only a one-dimensional simple gradient filter is applied while scanning the image from bottom to top and left to right for contour points that mark the transition between obstacle-free areas and obstacles (Figure 8). Applying the filter in one column is abandoned as soon as a contour point is detected or a given search height is reached that corresponds to a desired

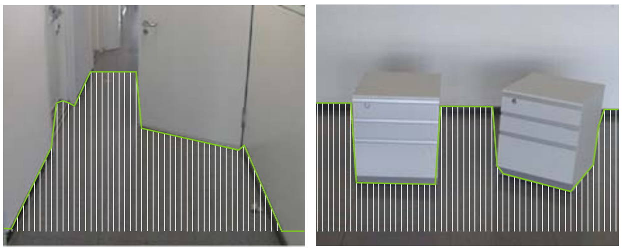

Figure 8: Gradient filtering along vertical search paths yields contour points that mark the transition between the floor and other objects. Left: typical corridor image with an open door as a possible obstacle. Right: two tables as possible docking objects.

look-ahead distance. Rows can be processed as well with a one-dimensional gradient-filter to ensure the detection of vertical edges, but only in the area being marked as obstacle-free after the column search. The remaining contour points are median-filtered to eliminate outliers and fitted to a polygon using an iterative end point fit algorithm. This yields a high-speed, yet robust, image segmentation into obstacle-free and obstacle-occupied regions. Another advantage of this algorithm is that its computation time is bounded to 100 ms in the worst case, i.e., when no obstacle is in front of the robot and both vertical and horizontal gradient filters are applied. A typical corridor image is processed within 30 ms, and if obstacles are present in the path of the robot, processing time automatically decreases even further, thus allowing a timely response.

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 that may be relevant for navigation, e.g., junctions, doors and docking stations. Upon object recognition, reference points and lines are identified (based on procedural and object data knowledge) and subsequently used for tracking.

Prerequisite for tracking objects is a reliable and fast image processing. For real-time vision the temporal coherence of scenes and the existence of physical objects should be reflected in the structure of the vision system, as proposed by [Graefe, Kuhnert 1988] (object-oriented dynamic vision). Moreover, even if image processing hardware and software have well advanced in recent years, it is still best to choose simple features and to verify their geometrical relationships during tracking. Edges and corners are the basis for our feature tracking algorithms. Edge points are fitted by a linear regression method to yield line parameters; corners are identified as line crossings. Search windows of moderate size around the chosen features of interest and adaptive thresholds are used to track them despite specular reflections and difficult illumination conditions, in general, at video field rate.

Kinesthesis

In organisms 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.

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.

3.4 Motor skills

Motor skills relate to 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., undercarriage, 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 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 human-like 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.

3.5 Sensorimotor Skills

Fixation-Based Skills

The ability to visually fixate objects or, more general, environmental points of interest is wide-spread among organisms. For our robots it is a good example of a sensorimotor skill. It first needs a sensor skill that continuously delivers the image coordinates of the point to be fixated with respect to a predefined reference 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 the reference point and the 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, obstacle avoidance and approaching workstations are examples for sensorimotor skills that were realized on this basis.

3.6 Skill-Based Calibration-Free Control

Complex robot behavior is achieved by concatenating the basic motor, sensor and sensorimotor skills described in the previous section. This has not only been demonstrated with our latest robot *HERMES*, but also with two other autonomous vehicles (*ATHENE I* and *II*) and a stationary manipulator. All robots were endowed from the start with certain basic skills, and were able to incrementally extend their knowledge and skills through learning.

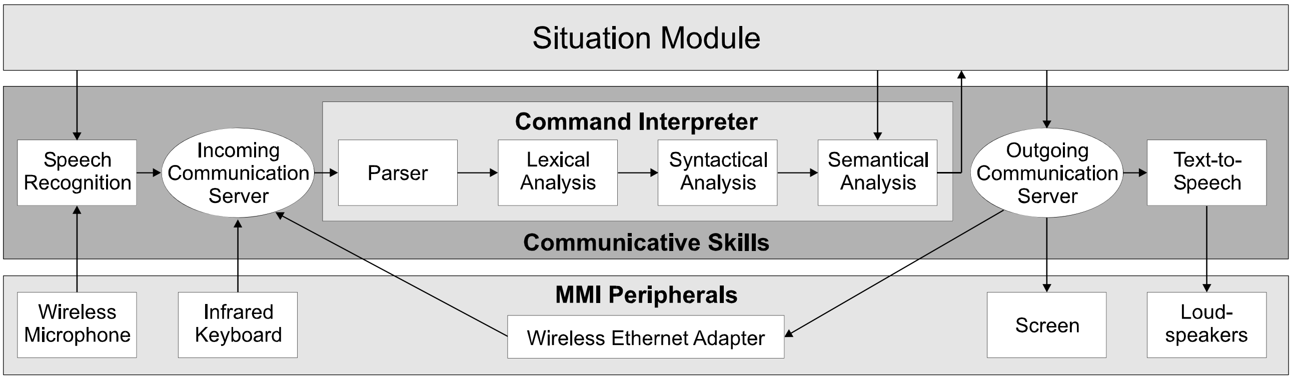

Figure 9: The man-machine interface. A communication server converts all incoming messages into an intermediate language. The translated messages are interpreted by the command interpreter and subsequently evaluated by the situation module, depending on the robot's actual situation. In turn, the robot's current situation directly influences the speech recognition and the semantic analysis to enhance recognition and interpretation. Outgoing messages are routed to the users via another communication server and appropriate communication channels (e.g., voice, graphics display or the Internet).

Calibration-free navigation

HERMES and two other autonomous vehicles (*ATHENE I* and *II*) were used as experimental platforms for navigation in networks of corridors in different buildings by concatenating simple vision- and odometry-based sensorimotor skills. It would have been difficult and time-consuming to measure all relevant kinematic parameters of all those robots, and especially of *HERMES* with its versatile undercarriage, with any degree of accuracy, and to design model-based controllers for all possible modes of operation. Our calibration-free approach makes it unnecessary.

In the case of wall-following it is based on the fact that a translation of a camera parallel to a straight guideline, e.g., the intersection of a wall and the floor, has no effect on the location of the line in the image. Simply speaking, the control strategy is to steer the vehicle in such a way that the motion of the guideline in the image that is caused by the vehicle motion is minimized. Moreover, the offset between the vehicle and the guideline, and the other control parameters, may easily be controlled at will (for details see [Graefe, Bischoff 2004]).

Calibration-free manipulation

A method for calibration-free object manipulation using a vision-guided stationary manipulator was first proposed by [Graefe 1995]. It is based on the idea that for grasping an object it suffices to make the object and the gripper coincide *in the images* of 2 cameras simultaneously. No knowledge of any camera or robot parameters is necessary, and changes in any of those parameters have no negative consequences. The method was later refined and implemented on arms like those of *HERMES* [Maryniak 2002].

3.7 Communication and Interaction

The communicative skills of *HERMES* are primarily based on natural language (English, French and German). It is used both to instruct the robot and to generate easy-to-understand messages for the user. Commands may be input via voice, keyboard or e-mail, while the robot may speak to the user, display its messages on a screen or send them by e-mail. Figure 9 shows the essence of the communication subsystem.

Command Interpretation

A command interpreter handles all user input. It consists of a parser, a lexical analysis, a syntactical analysis and a semantical analysis. The parser is fed by a text string that is provided by a communication server handling all incoming messages. Messages can come from the speech recognition module, the e-mail client or the keyboard. The parser separates the character string into a sequence of words and numbers. The words are given to the lexical analysis where each one is looked up in a dictionary to obtain its type.

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 and resolve pronouns such as “it”, “my” and “your” from the robot's situated knowledge in order to make the command complete.

Speech Recognition

We use a commercially available speech recognition engine (Lernout & Hauspie) for speaker-independent recognition of continuous speech. The speech recognition engine generates text strings equivalent to the ones that may be entered via the keyboard. An ordinary wireless microphone may be carried by the human for sending his utterances to the robot under truly adverse conditions. Normally it is sufficient, though, to attach the microphone to the robot's head.

Situated Context Switching

An important way to increase the robustness of the speech recognition system is the usage of so-called contexts that contain only those grammatical rules and word lists which 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

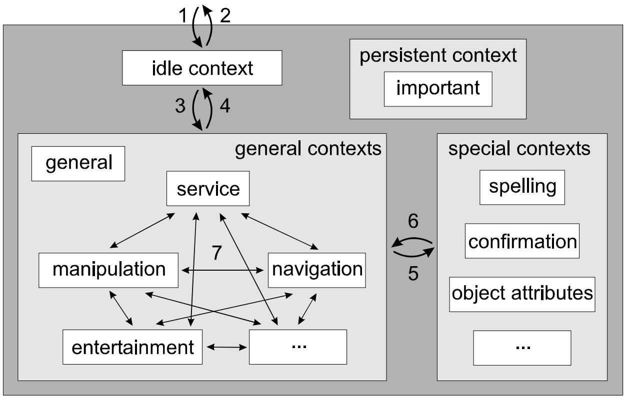

Figure 10: Visualization of the dependencies of the contexts implemented in *HERMES* for improving its communicative skills

it may expect from the user at a given moment, it can switch to a situation-specific context 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, many other words. By limiting the set of recognizable words or phrases to those which can actually be expected, the risk of recognition mistakes is reduced considerably.

Figure 10 visualizes some of the available contexts and how they are interconnected. At any stage in the dialogue a few words and sentences not related to the current context must be available to the user. They belong to a persistent context containing the most important commands that are recognized and executed with the highest priority. These commands 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 Moving" and "Halt" are used for disrupting the robot from its current task, "Cancel Command" and "Cancel Mission" will delete the currently executed command or mission from the command list. For an example of a dialogue see Figure 11, and for more details [Bischoff, Graefe 2002].

4 Experiment and Results

HERMES has proved both its intelligence and its dependability in many real-world experiments, including several public short term presentations, and in a long-term test. 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. In all our presentations we have experienced that the robot's anthropomorphic shape encourages people to interact with it in a natural and intuitive way. As presented in the preceding sections, *HERMES* possesses several other features that make it intrinsically more reliable and safer than other robots.

One of the most promising results of our experiments, especially the long-term test, 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, if 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.

Cooperative Learning Experiment:
In this experiment two tables have been placed in the field of view of the robot. HERMES ("H") finds itself in a state of waiting for a new instruction within the (default) service context while a person ("P") is approaching.
P: "Hello!"
H: "Hello. My name is *HERMES*. What is your name, please?"
P: "My name is Rainer."
HERMES cannot find the person's name in its database, and therefore, switches to the spelling context.
H: "Please spell your name!"
P: "Romeo Alpha India November Echo Romeo."
HERMES switches to the confirmation context.
H: "Your name is Rainer. Is this correct?"
P: "Yes!"
HERMES extends its database of persons and switches back to the service context.
H: "Hello Rainer. Pleased to meet you. What can I do for you?"
P: "Take this glass. Place it onto my table."
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 which table the user has in mind. HERMES has to ask the user to help it solving the current ambiguity.
H: "I have found two tables. Which one is yours?"
HERMES switches to the attributes context.
P: "The left one!"
HERMES switches back to the service context, approaches the left table and places the glass onto it. Then, it extends its database of objects, and marks the approached table as one of Rainer's tables.

Figure 11: Excerpt from a dialogue between a human and *HERMES* to place an object onto a table. In its course, *HERMES* learns more about its environment and stores this knowledge in several databases for later reference. It should be noted how often contexts are switched, depending on the robot's expectations, thus improving the speech recognition considerably.

The public presentations made us aware of the fact that a humanoid 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 the majority of cases *HERMES* behaved as expected, especially when it was operated by people who were used to work with it or, even better, had programmed it and knew all its features and flaws. However, even untrained persons not at all experienced or particularly interested in robotics, could easily operate *HERMES*, too.

Calibration-Free Navigation

For testing the calibration-free navigation concept vision skills were developed that yield the position of the guideline in the image and its distance to a reference point, *r*, in the image despite adverse conditions, such as specular reflections and difficult illumination conditions, in general, at video field rate (i.e., 50 Hz). Typically, prominent features in a corridor are visible at the intersections of the floor with the walls, as floors and walls usually have different colors. Such features may serve as guide lines.

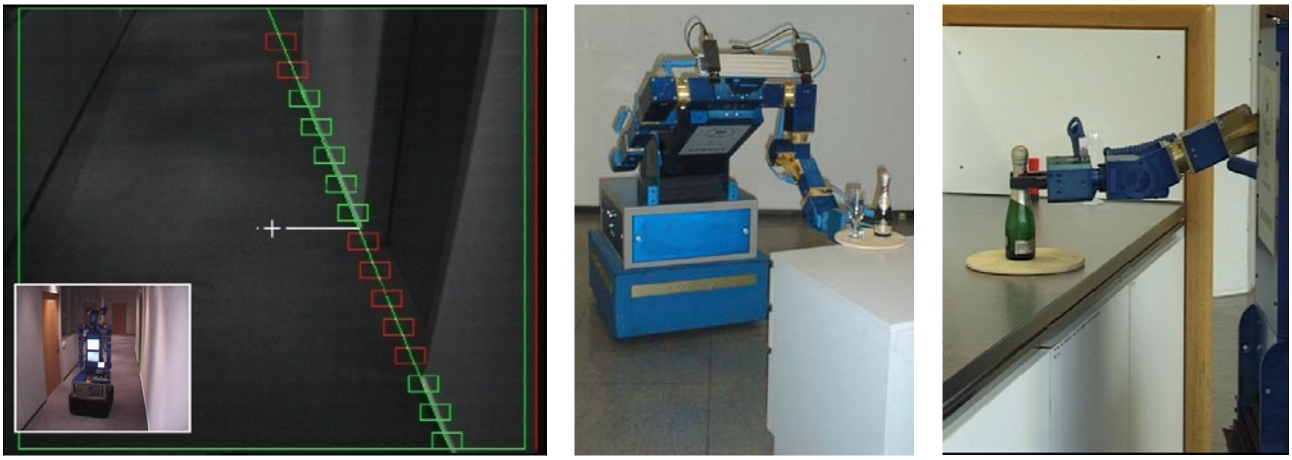

Figure 12: Experiments and Results. Left: Image taken by the *HERMES'* vision system during corridor navigation (tracking a guideline, indicating two forms of interruptions (door and junction). Middle and Right: Driving towards a table (docking) and placing a tray onto the table using the bendable body. Right: Placing a bottle onto a higher table (with upright body).

In the beginning of an actual driving experiment our mobile robots were placed somewhere in a corridor (Figure 12, left). Their exact orientation and position were unknown, but a section of the baseboard or other guideline was in the field of view. Corridor navigation was possible and proved to be remarkably robust. Unknown variations in camera pan and tilt angle were tolerated as well as replacement of the camera lenses by ones with different focal lengths. Even a Suematsu lens that, similar to the human eye, combines a wide field of view with a high resolution in the image center could be used without any modification to the control software. Based on these navigation skills map building and delivery missions were accomplished.

Manipulation

Manipulation experiments were successfully performed by interacting with people and in interaction with the environment. 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. When the human places the object between the robot's fingers, or takes the object, a touch event signals that the human has closed the kinematic chain and is willing to give or receive the 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 has touched something (Figure 12 middle and right).

Long-Term Experiments

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.

Demonstrating the robot's performance outside its normal laboratory environment, e.g., in television studios, subjects it to various kinds of stress. First of all, the robot might be exposed to rough handling during transportation, but even so, 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. *HERMES* has 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 the robot's capabilities, e.g., the robot is not yet able to act upon a directors command like a real actor (although sometimes expected!). But it is through 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 persons in the same hall make an incredible noise. It was an excellent environment for testing the robustness of *HERMES'* speech recognition system.

With respect to dependability the most significant evaluation was carried out during a long-term test where *HERMES* served in a museum, far away from its home laboratory, interacting with unknown humans and offering useful functionality to them for more than six months up to 12 hours per day. In a special exhibition the museum presented

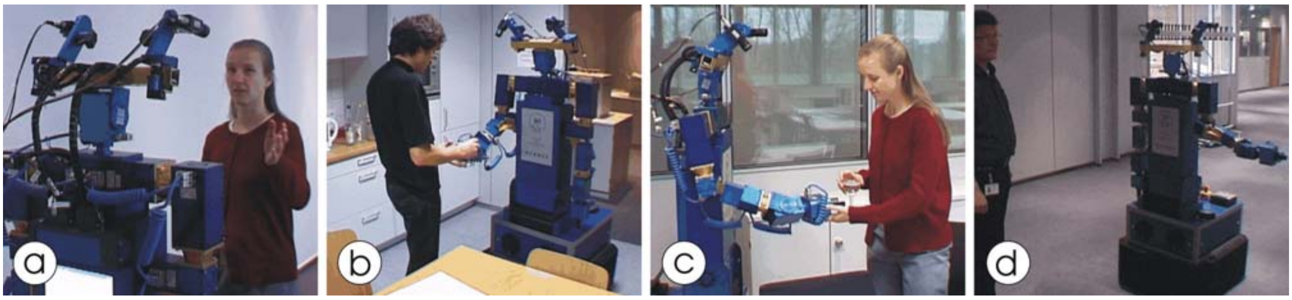

Figure 13: *HERMES* executing service tasks in the office environment of the Heinz Nixdorf MuseumsForum: (a) dialogue with an *a priori* unknown person, *HERMES* accepts the command to get a glass of water and to carry it to the person's office; (b) asking a person in the kitchen to hand over a glass 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 while gesturing with an arm and the eyes.

the current status 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 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 exposure time of the cameras. Visitors and personnel were most impressed by *HERMES'* plug-and-run capabilities and its human-friendly intuitive interaction interface. *HERMES* was able to chart the office area of the museum from scratch upon request and delivered services to *a priori* unknown persons (Figure 13 and also [Video A]). 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 personnel 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 from a bottle with water (a vision-guided sensorimotor skill), carrying it to a table and placing the glass onto it; and playing the visitors' favorite songs or telling jokes that were retrieved from the Web (Figure 14 and also [Video B]).

We found it interesting to observe how *HERMES*, actually just a laboratory prototype despite its designed-in dependability, survived the daily hard work far away from its "fathers" where no easy access to repair and maintenance was available, and how it got along with strangers and even with presenters who did not know much about robot technology. In fact, we were surprised ourselves that it performed so well. During 6 months of operation (lasting up to 18 hours a day during video recordings for documentation purposes) only one motor controller, one drive motor and one audio amplifier ceased to function, all of them commercially available and easily replaceable. According to the museum staff, *HERMES* was one of the few experimental robots that could regularly be demonstrated in action, and among them it was considered the most intelligent and most dependable one.

5 Summary and Conclusions

A bionic approach to the design of complex intelligent robots has been presented. *HERMES*, an experimental robot of anthropomorphic size and shape designed according to this approach, interacts dependably with people and their common living and working environment. It has shown robust and safe behavior with novice users, e.g., at trade fairs, television studios, at various demonstrations in our

Figure 14: *HERMES* performing at the special exhibition "Computer.Brain", instructed by natural language commands: 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 the visitors (here: waving with both arms, visitors wave back!) (c)

laboratory environment, and in a long-term experiment carried out in a museum and office environment.

The robot is basically constructed from readily available motor modules with standardized and viable mechanical and electrical interfaces. Due to its modular structure the robot is easy to maintain, which is essential for system dependability. A simple but powerful skill-based system architecture integrating visual, tactile and auditory sensing and various motor skills not relying on quantitatively exact models or accurate calibration are the basis for software dependability. Actively controlling the sensitivity of the CCD cameras makes the robot's vision system robust with respect to severe illumination contrast and varying lighting conditions. Consequently, safe navigation and manipulation even under uncontrolled and sometimes difficult lighting conditions are 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.

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 dependably 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. In contrast to today's robots that are mostly strong with respect to a single functionality, e.g., navigation or manipulation, our results illustrate that many functionalities can be integrated within one single robot through a unifying situation-oriented behavior-based system architecture. Our results 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 enable us to build dependable intelligent robots. We claim that these tests have to be carried out "here and now" by us robotics researchers, with the computing power and other resources available today. We hope that our results encourage others to include increasingly integration, dependability and long-term experiments in their research efforts.

References

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

Babcock, P. (1976): Webster's 3rd New International Dictionary of the English Language, G. & C. Merriam Company, Springfield, MA, 1976.

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

Bischoff, R.; Graefe. V. (2002): Dependable Multimodal Communication and Interaction with Robotic Assistants. 11th IEEE International Workshop on Robot and Human Interactive Communication. Berlin, pp. 300-305.

Bischoff, R.; Graefe, V. (2004): *HERMES* – a Versatile Personal Assistant Robot. Proc. IEEE – Special Issue on Human Interactive Robots for Psychological Enrichment, pp. 1759-1779.

Graefe, V. (1990): An Approach to Obstacle Recognition for Autonomous Mobile Robots. IEEE/RSJ International Workshop on Intelligent Robots and Systems, IROS '90. Tsuchiura, pp 151-158.

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.

Graefe, V. (2008): Calibration-Free Robots for Cost-Effective, Dependable and Robust Automation. IEEE International Conference on Automation and Logistics, ICAL 2008. Qingdao, China, pp 262-267.

Graefe, V.; Bischoff, R. (2004): Robot Navigation without Calibration. Proceedings IEEE International Workshop on Intelligent Robots and Systems (IROS '04). Sendai, pp. 627-632.

Graefe, V.; Kuhnert, K.-D. (1988): A High-speed Image Processing System Utilized in Autonomous Vehicle Guidance. Proceedings of the IAPR Workshop on Computer Vision. Tokyo, pp 10-13.

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

Maryniak, A. (2002): Vision-Based Manipulator Control of Calibration-Free Robots. Dissertation, Faculty of Aerospace Engineering, Bundeswehr University, Munich.

Videos: http://www.unibw.de/robotics/videos/

Video A: http://www.unibw.de/robotics/videos/ *HERMES* – a Versatile Plug-and-Run Service Robot

Video B: http://www.unibw.de/robotics/videos/ *HERMES* at the Exhibition