

# Integrating Vision, Touch and Natural Language in the Control of a Situation-Oriented Behavior-Based Humanoid Robot

Rainer Bischoff and Volker Graefe  
The Institute of Measurement Science  
Bundeswehr University Munich  
85577 Neubiberg, Germany

e-mail: {Rainer.Bischoff | Graefe}@unibw-muenchen.de

URL: <http://www.unibw-muenchen.de/campus/LRT6/welcome.html>

## Abstract

In the personal or service robotics domain a very close interaction between humans and robots is crucial. To facilitate such a close interaction we propose to design the robot according to an anthropomorphic model, to combine visual and tactile sensing and to base human-robot communication on natural language. Furthermore, we argue that integrating these key technologies into a robot requires a certain kind of a behavior-based system architecture that relies on an understanding of situations for the selection of the behaviors to be executed. A robot implementing this architecture allows a human interface to be realized that makes the robot appear intelligent and easy to communicate and interact with. Results of first experiments are reported.

## 1 Introduction

Man-machine interfaces are a key part of today's and future technical systems, especially in the personal or service robotics domain. Future personal or service robots will have to interact closely with humans, and specifically, with humans who are not robotics experts and often not even interested in robotics and other technical matters. It is considered unlikely that such persons will use interfaces like mouse or keyboard for instructing a service robot to perform specific tasks.

Studying how humans could possibly interact with such robots in a flexible manner leads to the conclusion that any communication and interaction should be as natural as possible. In other words: Communication with a robot should take place on a human-like level of abstraction and be conveyed via acoustic, visual or tactile communication channels that are employed by most humans and, thus, yield the highest possible degree of user-friendliness. If we consider, for instance, situations involving cooperative manipulation of objects by the robot and a human, both visual and tactile sensor information are indispensable: vision-based perception is necessary for navigation and for extracting relevant environmental information relating to distant objects, while tactile information augments the robot's perception for interactions with nearby persons or objects. To initiate this cooperative task and to exchange information during task performance, speech should be used as it is the preferred mode of human communication.

Many researchers are working towards the goal of truly human-friendly robots that have a number of different senses and can be safely operated and intuitively instructed. Although vision,

touch and natural language are major components in realizing human-friendly robots they have been studied rather independently because they constitute research areas in themselves. Therefore, not much work on autonomous robots with human-like manipulation capabilities operating in real environments and integrating *all three* components has been reported.

Jijo-2, a talking mobile robot without manipulators, is able to build a probabilistic map of its office environment by acquiring missing location information through conversational dialogues with people. A multi-microphone array has been developed to enhance the speech recognition rate under noisy real-world conditions [Matsui et al. 1996].

The dual-arm mobile robot KAMRO has been augmented with a natural language interface for specifying assembly tasks. Spatial relations between components to be assembled on the robot's workbench are identified by an overhead camera. Thus, a complete world model can be maintained. It may be used to complement or correct operator utterances that may be incomplete or inaccurate [Laengle et al. 1995].

The primary testbed of the collaborative research program "Situated Artificial Communicators" at the University of Bielefeld is a static assembly cell consisting of two cooperative robot arms equipped with multiple video cameras and force/torque sensors [Knoll et al. 1996]. Typed natural language is used to conduct the conversation between the machine and a human operator. A hybrid behavior-based system architecture has been developed that integrates language, perception and action. It handles situated references to actions and objects which could not be interpreted on the basis of the dialogue context alone [Milde et al. 1997].

The most sophisticated system up to date in terms of multimodal user interaction has been presented by [Thórisson 1998]. He created a prototype of a communicative "humanoid" that consists of a voice, a hand and a face which appear on a small monitor in front of the user. It is capable of real-time face-to-face dialogues with human users including various hand gestures, facial expressions, body language and meaningful utterances. Its underlying architecture is a model of human psychosocial dialogue skills that includes perception and interpretation of user behavior with respect to the current dialogue state.

With respect to existing systems and reviewing the corresponding literature we are able to identify three major problem areas that need to be solved for developing truly human-friendly personal robots:

1. Design and integration of the sensors and actuators necessary for enabling a robot to sense in, and act on, its environment in a human-like way.
2. Realization of a control structure that allows a robot to generate useful and goal-directed behaviors, based on a multitude of sensor inputs with regard to its actuation capabilities.
3. Development of communication and interaction behaviors to enable the robot to communicate intelligently and to display a user-friendly and cooperative attitude.

Our proposition to solve these problems is to build the robot according to an anthropomorphic model and to equip it with universally applicable sensor modalities, such as vision, touch and hearing, and building, thereupon, human-like sensorimotor and communication skills and behaviors that allow the robot to interact intelligently with persons and manipulate objects in its surroundings.

In the sequel, we explain these problems in more detail and present our solutions with respect to the state of the art. In section 3 we give an overview of the system architecture that allows us to integrate various sensor modalities in the control of our robot. Section 4 reveals some implementation details. Actual performance can be evaluated based on real-world experiments (section 5).

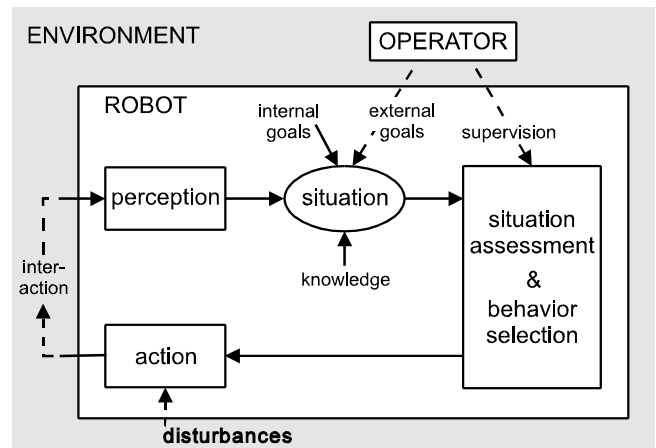
## 2 Major Problems and Solution Propositions

### 2.1 Designing Robots for Human Environments

Regarding the first problem, the robot design, it seems plausible to take a human as a design model. The best argument for such an approach is that a personal or service robot shall perform its tasks in environments where humans work and live, e.g., in apartments, offices, laboratories, restaurants, or hospitals. These environments are designed to meet special human characteristics and needs: the space a human requires, his working and vision height and his strength and the number of degrees of freedom available to manipulate objects [Bischoff 1998].

Instead of adapting the environment to the robot's needs it is a better approach to enable the robot to adapt itself to the environment. Shaping the robot according to an anthropomorphic model and equipping it with human-like sensor and motor skills will avoid subsequent and expensive changes of the infrastructure and make the robot, in principle, suited for any environment humans normally work and live in. Employing special purpose sensors, e.g., for distance measurement, may help in the short run with the development of *specific* systems; however, a more *generic* approach with sensors on the basis of vision, touch and hearing will in the long run lead to more cost-effective solutions and help to deploy future robots in massive numbers.

Another reason for an anthropomorphic design is that service robots will have to interact, and to communicate, with humans. If a robot has a humanoid form and exhibits human-like behavior, humans are able to interact with it in a more natural way. Also, movements of an anthropomorphic robot can more easily be predicted even by humans who are not interested in robot technology. Finally, humanoid size and shape of a robot can be advantageous for its representation of knowledge of the environment in such a way that it may easily be accessed by, and shared with, humans as a basis for communication. The closer the contact with humans the more humanoid the robot should be [Bischoff et al. 1998].



**Figure 1:** The role of situation as a key concept in the perception-action cycle of a situation-oriented behavior-based robot.

### 2.2 Controlling a Humanoid Service Robot

To solve the second problem, controlling a robot with many – partly redundant – degrees freedom in actuation and sensation, we propose to ground the system on a behavior-based architecture. Behavior-based system architectures are now generally accepted as an efficient basis for autonomous mobile robots [Arkin 1998]. Their main principle is the achievement of desired goals by activating an appropriate sequence, or combination, of behaviors that are selected from a repertoire of pre-defined behaviors.

The key problem in designing behavior-based robots is the question *how* to choose at each moment the most appropriate behavior. We propose to base this decision on a multitude of factors that we summarize under the term “situation”. The concept of “situation” here includes not only the objects in the environment and their states of motion, but also higher-level goals of both the human and the robot, overall tasks, and behavioral abilities of the robot.

Figure 1 illustrates how the concept of “situation” may be embedded in the perception-action cycle of a behavior-based robot. The figure also makes it clear that the “situation” on which the robot bases its behavior selection is, strictly speaking, only the robot’s internal image of the actual situation. Due to imperfect sensing or imperfect knowledge, this image may sometimes differ from the true situation, which will then result in a sub-optimal, or even grossly inappropriate, behavior of the robot. For further details see [Graefe, Bischoff 1997].

### 2.3 Communicating and Interacting with Service Robots

A user-friendly interface is a prime prerequisite for service robots that are aimed to help us in various activities in daily life. Their design and implementation is a two-fold problem: First, human and robot have to agree upon a suitable communication mode (or channel), second, utilizing the agreed-upon mode, communication and interaction have to be grounded on a common understanding or reference frame.

Since natural language is the easiest and most desirable mode of communication for a human it is desirable to integrate speech recognition and output into most service robots. Language can be used to instruct the robot with higher level goals or to intervene certain behaviors and modify their execution. However, to be accepted as cooperative partners, robots must not only have the ability to understand perfectly clear and complete commands, but they must also resolve ambiguities and complement

missing information that is inherent in human conversation. In doing so, an intelligent robot should pursue two approaches: One, it should use the current situation as a relevant context, and two, it may evoke additional information from the human through a dialogue.

Another important issue is how humans and robots can ground their communication and interaction on a common understanding or reference frame. In general, robots do not have the perceptual abilities of humans and, therefore, might not be able to detect the features of the environment a human would like to refer to during communication. In other words: While a human might want to describe his world in terms of salient features that he perceives or knows about, a robot has to use certain sensor readings and build special representations that, in general, do not conform with the way humans communicate and interact with each other.

However, communication can nevertheless be made possible by virtue of using common reference points in the environment and employing a common labeling for these points. For instance, language could be used to agree upon place or object names as a basis for cooperation. This does not necessarily require an understanding of the robot's representations of the world as long as the robot is able to refer to those agreed-upon points in its own terms when needed.

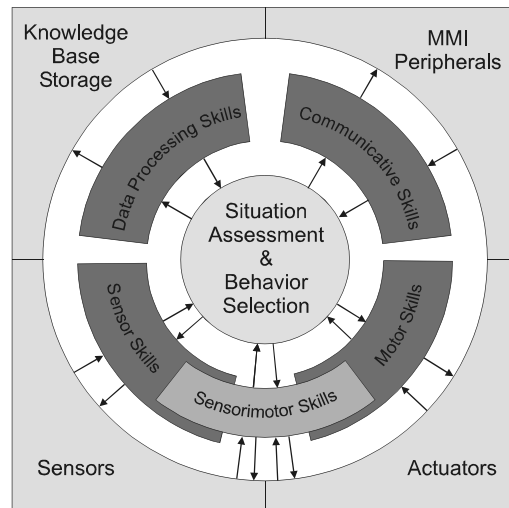
A situation-oriented approach, as proposed in the previous section, also seems to be adequate for solving this communication and interaction problem. Since man and machine are sensing and acting in a common environment, they will perceive their current situation in a similar way. If the robot is endowed with sufficiently powerful perceptual and cognitive abilities it may base its interpretation and execution of commands on that situation as perceived by both the robot and the human operator. It will then be able to accept orders of a similar form and abstraction level as those which a human would give another human and, thus, appear to be intelligent, cooperative and easy to communicate with.

### 3 System Architecture

Seamless integration of many different sensor modalities in a complex robot body calls for a unifying approach. We have developed a system architecture that allows integration of multiple sensor modalities and numerous actuators, as well as knowledge bases and a human-friendly interface, with access to all before-mentioned components on different levels of abstraction. In its core, the system is behavior-based, which is now generally accepted as an efficient basis for autonomous robots. 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.

#### 3.1 System Overview

Figure 2 shows the essence of the situation-oriented behavior-based robot architecture as we propose it. The situation module



**Figure 2:** System architecture of a personal robot based on the concepts of situation, behavior and skills.

(situation assessment & behavior selection) acts as the core of the whole systems and is interfaced via “skills” in a bidirectional way with all other hardware components – sensors, actuators, and knowledge base storage and MMI peripherals (man-machine and machine-machine interface peripherals).

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

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

#### 3.2 Skills and Behaviors

According to [Babcock 1976] a skill is the “knowledge of the means or methods of accomplishing a task” or the “ability to use one’s knowledge effectively and readily in execution or performance” or the “dexterity or coordination especially in the execution of learned physical tasks”.

According to psychological literature [Proctor, Dutta 1995], skills have several characteristics in terms of which they can be defined: First, a skill is acquired through practice or training, i.e., it is learned, but not innate. Second, skill develops in response to some demand imposed by the task environment on the organism. Third, skill is said to have been acquired when the behavior is highly integrated and well organized. Finally, cognitive demands are reduced as skill is acquired, freeing limited mental resources for other activities.

We derive our definition of skill from these characteristics, except that, for practical reasons, low-level skills need not to be learned (from scratch) but are pre-programmed and in-built (e.g., basic motor and sensor skills such as moving a joint or getting joint angle, see below for some more examples). One could refer to these low-level skills as the robot’s basic *abilities* upon which high-level skills may evolve in response to some external demand. High-level skills may also be pre-programmed, but may contain a learning component, that allows their continuous improvement over life-time. To summarize, we define a skill as a goal-directed, well organized behavior that is in-built, *can be* acquired and improved through learning, and is performed with economy of effort.

In this sense a skill can be the basis for an observable behavior. According to [Babcock 1976] a behavior is an “observable

activity when measurable in terms of quantifiable effects on the environment whether arising from internal or external stimulus” or “anything that an organism does involving action and response to stimulation” or “the response of an individual, group, or species to its environment”. Thus, behavior is more general than skill in a sense that it also includes purely reactive responses to external or internal stimulations, e.g., reflexes such as hard-wired emergency stopping, or may be evoked by emotions such as anger or happiness.

In general, most skills involve the entire information processing system. However, at a gross level, they can be divided into those with primarily perceptual, motor or cognitive components. Regarding the proposed system architecture, cognition takes place at the situation module. Selecting a specific behavior or skill is, therefore, an expression of a cognitive process.

All hardware components are accessed by the system via skills which can be classified into five categories:

**Motor skills:** Motor skills are simple movements of the robot’s actuators. They can be arbitrarily combined to yield a basis for more complex control commands. Encapsulating the access to groups of actuators, that form robot parts, such as wheelbase, arms, body and head, leads to a simple interface structure, and allows an easy generation of pre-programmed motion patterns. Examples for motor skills are: *move single module to a certain position*, *set velocity of a single module*, *open/close grippers*, *move arm to a certain position*, *move arm relative to the current position*, *move robot into initialization position*, etc.

**Sensor skills:** Sensor skills encapsulate the access to one or more sensors, and provide the situation module with proprioceptive or exteroceptive data. Angle encoders, motor current converters and battery gauges would be useful proprioceptive sensors giving feedback about the robots internal state. Visual, tactile and olfactory sensor modalities could provide useful data about the robot’s environment. Examples for sensor skills are: *get position of module*, *get position of arm*, *get grasping force*, *detect docking station*, *track person*, *detect touch* etc.

**Sensorimotor skills:** A sensorimotor skill combines both sensor and motor skills to yield sensor-guided robot motions. Vision-guided motions skills are, e.g., *find object* (by actively controlling the camera head according to a search pattern and searching for the object in the images), *fixate/track object*, *visually servo the gripper to a goal location*, etc. Tactile and force/torque-guided motion skills are, e.g., *place object onto table*, *maintain contact on a surface* (with specified force during motion), *take object from/give object to a person* (moving the arm towards the person and subsequently detecting contact), *shaking hands with person* (conforming to person’s motions), etc.

**Communicative skills:** Communication takes place via various different channels (acoustic, optical, tactile) and requires a variety of communicative skills to pre-process user input and to generate a valuable feedback for the user according to the current situation and the given application scenario. Communicative skills depend on various sensor skills to receive the required information, e.g., hearing skills for understanding speech, vision skills for recognizing gestures or mimics, and even tactile skills, if braille reading were to be performed. Sending information to a receiver can be as simple as switching a light on or off or as complex as producing speech. Examples for communicative skills are: *convert natural language into text* (and vice versa), *display status information*, etc.

**Data processing skills:** 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 its data bases. They can also provide means of more complex data processing, for example: *plan the shortest route* (through a given network of corridors), *provide next task description*, etc.

## 4 Implementation

### 4.1 Design and Realization of HERMES

In designing our humanoid experimental robot named *HERMES* we placed great emphasis on modularity and extensibility [Bischoff 1998]. All drives are realized as modules with compatible mechanical and electrical interfaces; each drive module contains a motor-transmission combination, power electronics, sensors, a micro-controller, and a communication interface (CAN bus) that connects them with the main computer.

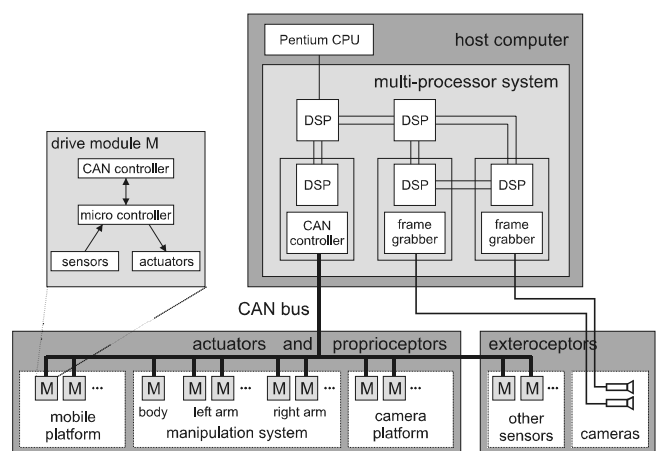
*HERMES*’ driving mechanism is omnidirectional. The manipulator system consists of two articulated arms with 6 degrees of freedom each on a body that can bend forward and backward. Currently each arm is equipped with a two-finger gripper that is sufficient for basic manipulation experiments.

Main sensors are two video cameras on a pan-tilt platform. Numerous proprioceptors such as angle encoders, current converters or temperature sensors, are integrated in the motor modules; additional sensors may be connected via available interfaces. A radio Ethernet interface allows to control the robot remotely. Separate batteries for the motors and the information processing system allow a continuous operation of the robot for several hours without recharging.

### 4.2 Control

A hierarchical multi-processor system is used for information processing and robot control. The control and monitoring of the individual drive modules is performed by the sensors and controllers embedded in each module. The main computer is a network of digital signal processors (DSP, TMS 320C40) embedded in a standard industrial PC. Sensor data processing (including vision), situation recognition, behavior selection and high-level motion control are performed by the DSPs, while the PC provides data storage and the human interface (Figure 3).

Overall control is realized as a finite state machine that is capable of responding to events and interrupts and that can handle



**Figure 3:** Modular and adaptable hardware architecture for information processing and robot control.

and exchange messages between the different processors involved. After powering up the robot finds itself in the state "Waiting for next mission description". A mission description is provided as a text file that may be either loaded from a disk or received via e-mail or entered via keyboard. It consists of an arbitrary number of single commands or embedded mission descriptions that let the robot perform a required task. All commands are written in natural language and passed to a parser and an interpreter. If a command cannot be understood, 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.

In the current implementation commands have to be typed and the robot's responses are written to a display. Integration of speech recognition and output is well under way, so that in the end both written and spoken instructions will be understood.

All of the example skills listed in section 3.2 were actually implemented, except for those requiring force/torque sensing [Bischoff 1999]. Motion 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 are implemented on those DSPs that have direct access to digitized sensor data, especially digitized images.

#### 4.3 Knowledge Representation

For a situation-oriented behavior-based mobile robot, that has to pursue certain goals, in principle, three different types of knowledge bases are needed: one, a knowledge base containing information about the static characteristics of the environment, two, knowledge about the mission to accomplish and general goals to pursue, and three, knowledge on how a present situation could be changed to the desired one by executing appropriate behaviors. An intelligent robot as we have it in mind should be able to communicate its knowledge to humans and to learn and extend its knowledge by interacting with the world.

**Attributed topological map:** The static characteristics of the environment are stored in a topological map that is augmented by certain attributes [Graefe, Bischoff 1997]. Intersections and junctions as, well as task-relevant locations, such as docking stations, are modeled as points, and passageways connecting such locations as pairs of anti-parallel directed paths. Points and paths carry attribute lists that contain relevant information, e.g., approximate information on geometrical quantities (e.g., lengths of, and angles between, corridors), as well as descriptions of visible landmarks and hints to certain behaviors that are related to specific locations, e.g., "go slowly" or "keep right".

**Mission description:** Depending on his preferences and on the abilities of the robot the user may define the robot's mission in more or less detail. A mission description may either consist of a detailed list of actions (e.g., elementary behaviors) that are to be executed sequentially by the robot, or only of a single command if the user has enough confidence in the robot's planning abilities.

Obviously, this type of communication with a robot on the basis of high-level task descriptions resembles instructions that might also be given to a person. However, only a mission description that refers directly to the elementary in-built behaviors and skills of the robot can eventually be executed. If a mission description consists originally of other items than these elementary

behaviors the robot's data processing skills in combination with its cognitive skills must convert it into a sequence of in-built behaviors which may then be executed (see section 5).

**Behavioral knowledge:** The knowledge necessary for situation recognition and robot control is represented partly in procedural form within the various algorithms, and partly in declarative form in data bases, e.g., how to break down complex skills or behaviors in more simpler ones that can be actually executed, and how objects can be recognized and manipulated.

## 5 Experiments

We conducted a number of real-world experiments with the humanoid robot *HERMES* to evaluate the concept presented in the preceding sections. An example that may serve to show the potential of the concept, but also the limitations of the current implementation, is depicted in Figure 4. The corresponding dialogue with a human user is reprinted in the sequel:

**Human: "Go to the secretariat and wait for instructions."**

Being in a state to wait at its home base for the next mission to accomplish or command to execute, *HERMES*' situation changes as it receives this new command. After parsing it *HERMES* plans a route from its current location to the secretariat. The route is planned by the appropriate data processing skill (Dijkstra shortest path algorithm) in terms of vision-guided navigation behaviors, e.g., leave home base, turn right, turn left at the second intersection, turn right, stop at the second door to the right (no coordinates are used). Then, the planned behavior sequence is executed, and if nothing unplanned happens (e.g., door locked) the robot reaches a position in front of the secretariat where it reports to the user.

**HERMES: "Hello! I am waiting for further instructions!"**

**Human: "Take over tray."**

From its data base the robot knows what a tray looks like and how it has to be grasped so that the objects on it will not fall down. Since the robot is not skilled enough to perceive the current pose of the tray it has to bring its arm into a configuration where the human can easily hand over the tray.

**HERMES: "Place tray into my left hand, please!"**

Now *HERMES* waits for a touch event that it can recognize by its proprioceptive sensors. A touch event is detected as an involuntary unexpected movement of one or more of its joints that most probably results from external forces. As *HERMES*' encoders yield an angle resolution of better than  $0.05^\circ$  and can be sampled at a rate of up to 1 kHz, very small external forces can be detected and, thus, utilized to signal a touch event. After *HERMES* has successfully grasped the tray, it again waits for further instructions.

**Human: "Place it on the table in the kitchen, then go home."**

This sentence invokes a number of different behaviors. First the robot has to find out what object the user has in mind when referring to "it", then it has to check whether it knows the location (the table in the kitchen) and finally has to plan its way to the location. If all parameters are known, the robot will break down the instruction into elementary behaviors. To verify if the robot has well understood what the user requested it repeats what it has understood and asks for confirmation.

**HERMES: "I will go to the kitchen, then I will dock to the table, then I will place the tray onto the table, then I will go**

home. Do you want me to do these tasks?"

Human: "Yes, please!"

By executing the before-mentioned behaviors *HERMES* is fulfilling the tasks as required. A path is planned as a sequence of elementary hallway navigation behaviors towards the kitchen. Since the table within the kitchen might have been moved to a different location, docking to the table is performed using a visual servoing method [Bischoff, Graefe 1998]. The behavior of placing an object onto another object is again performed utilizing a sense of touch (similar to the one described above). Thereafter, *HERMES* returns to its home base location.

This dialogue is a representative example of some of the capabilities of the robot. Natural language recognition is currently restricted to written language and to imperative sentences that have a relatively simple structure. Nevertheless, the robot shows already fairly cooperative behavior.

## 6 Conclusions and Outlook

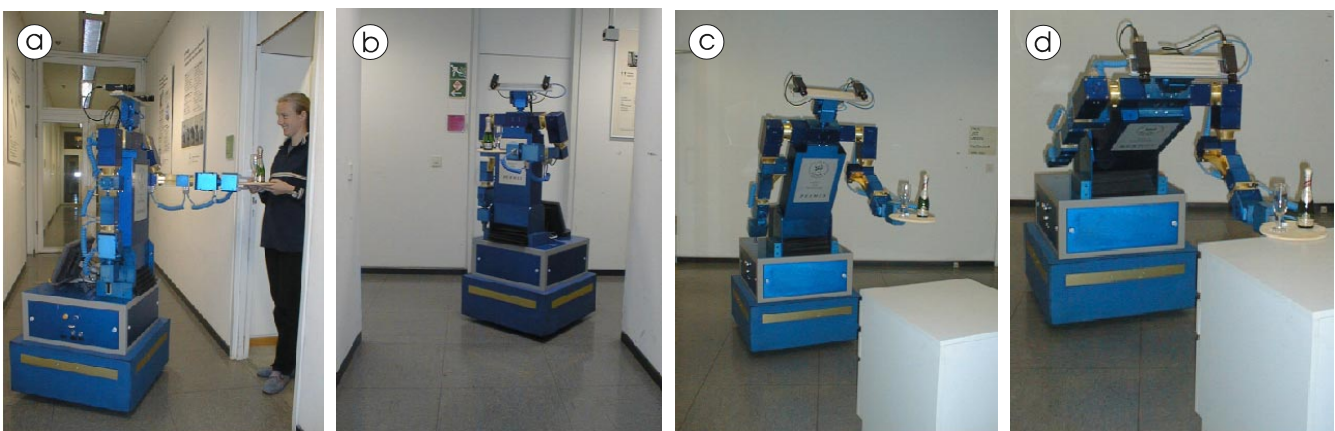
By integrating various sensor modalities including vision, touch and hearing a robot may be designed that displays intelligence and cooperativeness in its behavior and communicates in a user-friendly way. This was demonstrated in experiments with a complex robot designed according to an anthropomorphic model. A special kind of behavior-based system architecture has been proposed to control the robot. Its main idea is to select and coordinate the behaviors based on an assessment of the situation being perceived by both the human operator and the robot at a particular moment. This concept places high demands on the robot's sensing and information processing, as it requires the robot to perceive situations and to assess them in real time. A network of microcontrollers and digital signal processors embedded in a single PC in combination with the newly introduced concept of skills for organizing and distributing the execution of behaviors efficiently among the processors, is able to meet these demands. Skills can be improved through learning during the life-time of the robot.

Due to the innate characteristics of the situation-oriented behavior-based approach the robot is able to cooperate with a human and to accept orders that would be given to a human in a similar way. Just as a human, the robot does not expect its orders to be expressed in terms of coordinates or exact distances; place names and semi-qualitative references, such as "at the next intersection" suffice.

To improve the human-friendliness of the robot further, speech recognition and speech output capabilities will be added, based on the existing grammar, and the grippers will be equipped with tactile sensors.

## References

- Arkin, R. C. (1998).** Behavior-Based Robotics. MIT Press, Cambridge, MA, 1998.
- Babcock, P. (1976):** Webster's Third New International Dictionary of the English Language, G. & C. Merriam Company, Springfield, MA, 1976.
- Bischoff, R. (1998).** Design Concept and Realization of the Humanoid Service Robot *HERMES*. In A. Zelinsky (ed.): Field and Service Robotics. Springer, London 1998, pp. 485-492.
- Bischoff, R. (1999).** Advances in the development of the humanoid service robot *HERMES*. Second International Conference on Field and Service Robotics. Pittsburgh, PA, Aug. 1999, to appear.
- Bischoff, R.; Graefe, V. (1998).** Machine Vision for Intelligent Robots. IAPR Workshop on Machine Vision Applications. Maku-hari/Tokyo, November 1998, pp. 167-176.
- Bischoff, R.; Graefe, V.; Graßnick, R. (1998).** Serviceroboter der Zukunft. Spektrum der Wissenschaft. Dossier (4/1998): Roboter erobern den Alltag. Heidelberg, Oct. 1998, pp. 66-71 (in German).
- Graefe, V.; Bischoff, R. (1997).** A Human Interface for an Intelligent Mobile Robot. 6th IEEE International Workshop on Robot and Human Communication, Sendai, Japan, Sept. 1997, pp. 194-197.
- Knoll, A.; Hildebrandt, B.; Zhang, J. (1996).** Instructing Cooperating Assembly Robots through Situated Dialogs in Natural Language. Research Report 96-01. Uni Bielefeld, Techn. Fakultät.
- Laengle, T.; Lueth, T. C.; Herzog, G.; Stopp, E.; Kamstrup, G. (1995).** KANTRA - a natural language interface for intelligent robots. In U. Rembold and R. Dillmann, eds., Intelligent Autonomous Systems IAS-4, pp. 365--372. IOS Press, 1995.
- Matsui, T.; Asoh, H.; Asano, F. (1997).** Map Learning of an Office Conversant Mobile Robot, Jijo-2, by Dialogue-Guided Navigation. Proceedings of the International Conference on Field and Service Robotics. Canberra, Australia, December 1997, pp. 230-235.
- Milde, J.-T.; Peters, K.; Strippgen, S. (1997).** Situated Communication with Robots. First International Workshop on Human-Computer Conversation. Bellagio, Italy, July 1997.
- Proctor, R. W.; Dutta, A. (1995).** Skill Acquisition and Human Performance. In: Bourne, L. E. (ed.). Advanced Psychology Texts. SAGE Publications, Thousand Oaks, 1995.
- Thórisson, K. R. (1999).** A Mind Model for Multimodal Communicative Creatures and Humanoids. Int. Journal of Applied Artificial Intelligence, Special Issue on Animated Interface Agents, to appear.



**Figure 4:** *HERMES* performing a simple service task upon user request: a) receiving a tray (with a bottle and a glass) from a user; b) navigating in a network of hallways towards the commanded goal location; c) approaching a table at the goal location and d) placing the tray onto it.