11th IEEE International Workshop on Robot and Human Interactive Communication. Berlin, September 2002

Dependable Multimodal Communication and Interaction with Robotic Assistants

Rainer Bischoff and Volker Graefe
Bundeswehr University Munich, The Institute of Measurement Science
85577 Neubiberg, Germany
E-Mail: Bischoff@ieee.org, Graefe@unibw-muenchen.de
http://www.unibw-muenchen.de/hermes

Abstract

To advance research in the field of multimodal human-robot communication we designed and built the humanoid robot HERMES. Equipped with an omnidirectional undercarriage and two manipulator arms it combines visual, kinesthetic, tactile, and auditory sensing with natural spoken language input and output and body expressions for natural communication and interaction with humans. HERMES was successfully tested in an extended six-month experiment in a museum where only naive users interacted with the robot. They chatted with HERMES in several languages and requested various services. Multimodal communication and an understanding of the current situation by the robot turned out to be the key to success.

1 Introduction

Human-robot interaction. A robotic assistant equipped with an ideal communication and interaction system would allow a human to ask for what he needs in any way he chooses. Human-robot interaction would be similar to the everyday interaction among humans. Especially when robots and non-expert humans have to interact and collaborate, a human-like communication and interaction is desirable.

Human conversation is usually multimodal. Multimodality enhances the richness of the communication and interaction and allows more complex information to be conveyed than is possible with a single or two modalities. Consequently, implementing most of the human senses and communication channels is a prime prerequisite for cooperative and user-friendly service robots. Ideally, they should, in addition to communication via spoken language, generate and understand gestures, body and facial expressions and touch events, to fully support a truly human-friendly interaction.

Multimodal conversations with autonomous robotic assistants are in principle not predictable and rely on mixed dialogue initiatives. It might be that the robot needs to urgently inform its human partner about some problems (e.g., batteries running low) or to remind him of his agenda (e.g., to take some medicine). On the other hand, the user needs to be able to request the robot's services at any time in an easy (and therefore unpredictable) way.

Figure 1: *HERMES* describing the way to a location of interest via voice and gestures

In their conversations humans often refer to past and current utterances and to perceived environmental features. To understand instructions in natural language, the robot must, therefore, be able to interpret them in a context- and situation-dependent way. To converse naturally, the robot, too, must formulate its own utterances in a similarly context- and situation-dependent way. The variety of possible situations the robot might find itself in, and the number of expressions the human might use, make this a real challenge for current robot technology, in particular for perception, situation recognition and speech understanding.

Related work. Many researchers are working towards the goal of truly human-friendly robots that have a number of different senses and can be safely operated and intuitively instructed. Although vision, touch and natural language processing in combination are the key to realize human-friendly robot interfaces, they have mostly been studied rather independently because they constitute research areas in themselves. Therefore, hardly any work on real robots operating in real environments and integrating *all three* components has been reported. The most advanced systems in terms of integrating vision-based scene understanding and language recognition in the robotic assembly domain were presented by [Laengle et al. 1995] and [Knoll et al. 1996], and in the mobile robot domain by [Torrance 1994] and [Matsui et al. 1997]. [Lemon et al. 2001] used an autonomous helicopter for their man-machine experiments based on a spoken-dialogue system combined with pointing at map features.

An experimental robot. In contrast to those researchers who study human-robot communication in simulations or using only part of a robot, e.g., a head, we prefer to work with real robots performing real services in the real world. We, therefore, designed and built a humanoid mobile robot, *HERMES*, and used it for our experiments (Figure 1). It is designed according to a situation-oriented skill- and behavior-based paradigm [Bischoff, Graefe 1999]. A central situation recognition module controls all perception, action and communication; it integrates sensory data, maintains knowledge bases, and selects dynamically the behaviors to be executed in each situation. *HERMES* can navigate in buildings and build maps, manipulate objects, perform errands, entertain people and give information. To interact with humans it uses mainly spoken language, but also gestures and its tactile and kinesthetic senses. It engages in true dialogues with multiple commands, statements, questions and answers (multiple turns), where both dialogue partners can take the initiative. If one or more pieces of information are missing in an instruction or question the robot asks questions to fill in the missing pieces. The underlying techniques are described in the sequel.

2 The Realized Communication System

The communicative skills of *HERMES* are primarily based on natural language. 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.

To enable natural language understanding with limited computational resources, a special grammar was designed. Examples for command sentences are object and action-oriented instructions such as “Go to the kitchen!”, or “Grasp the small ball!”. Directive instructions such as “Turn around!” or more complex commands like “Turn left at the next intersection!” are supported as well. Intervening commands that do not contain a command verb are partly supported, e.g., “faster” (instead of “move faster”). In this case these adverbs are treated like single command words. Questions that start with specific key words are allowed as well, e.g., “What”, “Where” and “How”. Numerous questions relating to facts stored in the in-built data bases or available through the World Wide Web can be answered by the robot, e.g., “What can you do?”, “Where are you?”, “How do I get to Rainer's office?”, “What is Rainer's phone number?”, “What is the current weather report for Munich?”, etc. The fixed syntax presently does not allow an arbitrary reordering of parts of the sentences, e.g., “Take the glass, the big one” or “The glass over there, please take it”.

Command Interpretation. A command interpreter handles all user input. It consists of a parser, a lexical analysis, a syntactical analysis and a semantical analysis (Figure 2). 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, using space, tabulator and punctuation characters as delimiters. The words are given to the lexical analysis where each one is looked up in a dictionary to obtain its type. Possible types are command verbs (e.g., go, take, place), locations (e.g., office, kitchen, workshop), prepositions (e.g., to, on, onto, in, into), objects (e.g., ball, table, pen) and fill words (e.g., please), just to name a few. Character strings enclosed in quotation marks are treated as one part of a sentence of type “text string”.

The following syntactical analysis tries to identify the structure of the sentence by comparing the list of types with a list of prototype command sentences that includes all the commands the robot is able to understand. If the comparison is successful the semantical analysis will eventually provide missing words 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. We provided an ordinary wireless microphone to be used 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.

To render the speech recognition more robust, larger word classes such as [object] have been split into several classes, e.g., [object_to_be_manipulated] and [object_used_-for_navigation] which are now used as specific arguments of the command words TAKE or GRASP and MOVE or GO. The fewer the number of words per class and the stricter the syntax, the better the results of the speech recognition will be because fewer hypotheses have to be verified. Also, meaningful results are produced even under noisy conditions. Another advantage is that the recognition of a specific gram-

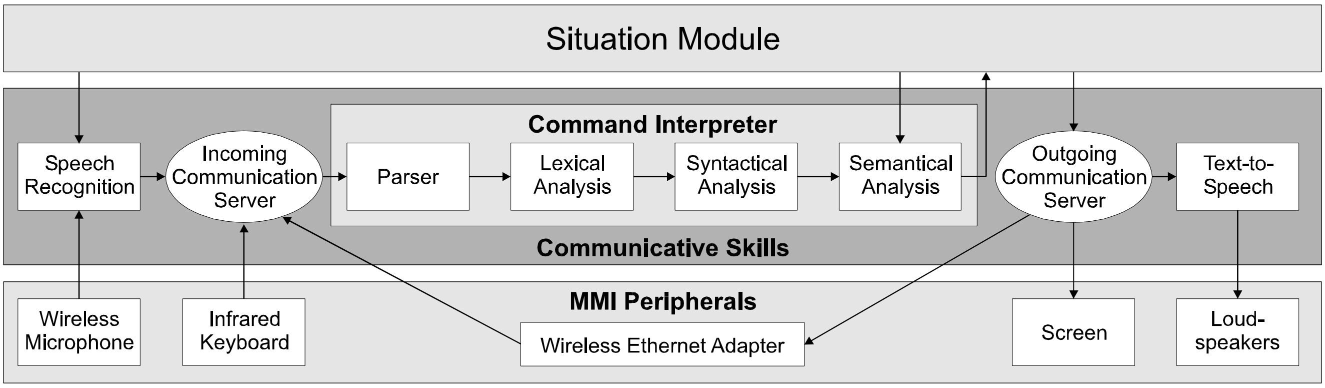

Figure 2: The man-machine interface. A communication server handles all incoming messages. They are interpreted by the command interpreter and subsequently handled by the situation module, depending on the robot's actual situation. In turn, the robot's current situation directly influences the speech recognition and the 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)

matical structure can be exploited to detect out-of-vocabulary words. To incorporate these words into the robot's vocabulary, a sub-dialogue is initiated that asks the user to spell the unknown word. Once the word has been successfully spelled it can be transcribed into phonemes which in turn allows its addition to the robot's language database.

Situated Context Switching. Another important way to increase the robustness of the speech recognition system has been the usage of so-called contexts that contain only those grammatical rules and word lists that are needed for a particular situation. Most parts of robot-human dialogues are situated and built around robot-environment or robot-human interactions, a fact which may be exploited to enhance the reliability and speed of the recognition process. When the robot knows what kind of answers it may expect from the user at a given moment, it can switch to a situation-specific 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 3 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.

Switching the robot on will enable (Figure 3, (1)) the "idle" context for command interpretation. This context contains various greeting formulas to address the robot while it is not expecting any user input, e.g., when it has nothing to do or is autonomously executing a task that does not require user inputs. This context has been designed to be very insensitive to noise, i.e., command sentences are limited and need a high recognition rate to be accepted. Speech recognition can be disabled completely (2) by the command "Stop Listening". (It may then be re-enabled manually.)

Once a greeting formula has been recognized, the robot switches from the idle context to a more general class of contexts (3) (For an example see also Figure 4.). In this class of contexts a "basic" context and a dialogue-dependent context (service, manipulation, navigation, entertainment, ...) are activated in addition to the before-mentioned persistent context with important commands. The "basic" context contains all types of questions to inquire the robot's databases or situated knowledge, e.g.: "What time is it?", "How do I get to Rainer's office?" etc. In addition, it contains commands and questions for switching explicitly or implicitly from one context to the next (7). For instance, "Could you do me a favor?" would implicitly activate the "service" context, "Entertain me, please!" would activate the "entertainment" context whereas "Switch to <context>" would explicitly activate a user specified context ("<context>" stands for the name of any known context). The robot automatically returns to the idle context when a dialogue is over (4).

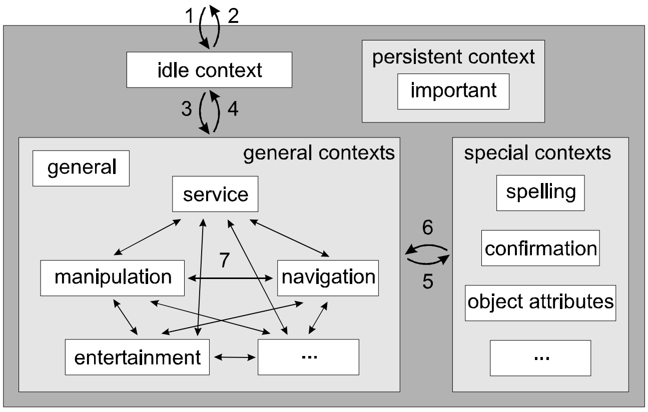

Figure 3: Visualization of the dependencies of the contexts implemented in *HERMES* for improving its communicative skills (see text for further explanations)

The general contexts were designed around standard human dialogues in the domains of service, manipulation, navigation and entertainment. Others may be added with ease. Each context contains multiple phrases and many words to allow complex robot control for, e.g., fetch and carry tasks, environmental exploration or fun applications. Their design allows even novice users to intuitively instruct the robot to perform a certain task. Although initially the user is allowed to take the initiative in the dialogue, it is the robot that is taking over once it has grasped the user's utterances. If the robot has not fully understood a command, or if some command parameters are missing, it asks a question to fill in the missing information (Figure 4, H's third utterance).

Depending on the prevailing situation, the type of dialogue conducted and the dialogue state, various "special contexts" may be activated (5). These special contexts further enhance the speech recognition in frequently occurring dialogue situations, such as confirmation or spelling. Special contexts can be very simple, e.g., allowing only "Yes" or "No" when *HERMES* expects such an answer, or more complex. For example, to teach the robot object and place names, a spelling context was defined that mainly consists of a spelling alphabet. Since spelling alphabets are optimized for ease of use by humans in noisy environments, such as aircraft, it should be well suited for robotic applications, too (except for the fact that it is not intuitive for naive users).

Special contexts are only activated temporarily for a specific part of the full dialogue. They are deactivated as soon as the required information has been obtained; then the previously active context is activated again (6).

Multiple Language Communication. The communication servers map the available natural languages into an internal robot language and vice versa, thus realizing multilingual speech recognition and speech output. All contexts were defined in the three languages English, French and German. It is possible to change the language within an ongoing dialogue while the actual context (as selected by the situation module) is preserved.

Dialogue structuring. One approach to make speech communication more robust and dependable, is to have the robot engage the human in a dialogue instead of having the human engage the robot. This makes the human’s speech much more predictable and leads to better recognition and language processing. On the negative side, this approach limits the means of interaction, probably making the interaction less efficient. Also, the human might be asked to provide more information than what is actually necessary in a specific situation, which would be tiresome.

Well-designed prompts are critical for efficient dialogues. They serve two purposes: (1) as a cue when it is the user’s turn to speak, (2) as an indication of what may be spoken. While explicit prompts (or directive prompts) specify the exact words or phrases the user should say, implicit prompts are open-ended, i.e., they do not list possible responses. Implicit prompts can provide a more natural communication for the user, but allow more room for misunderstandings, whereas explicit prompts are useful in constraining user responses (see Figure 4 for examples).

3 Multi-Modality Supporting Communicative Goals

From their day-to-day experience humans are very skilled at interpreting human behavior. Therefore, it makes sense to have the robot conform to those social norms that have evolved to facilitate human interaction. Consequently, it is not enough for a robotic assistant to use speech as the only means to communicate with humans. Other communication modes should be employed as well, such as facial expressions, touch events, gestures and poses, to make the communication richer and more effective. An anthropomorphic shape is certainly advantageous, e.g., to exchange objects with humans and to point and look in directions of interest.

Although humanoid in its appearance, *HERMES* does not possess an animated face to attract people or to indicate

Utterances / Actions (P: Human; H: *HERMES*)	**Activated Context**	**Comments**
P: “Hello!”	idle	robot is disrupting its current task
H :“Hello! What can I do for you?”	service	normal prompt (the robot expects a moderately experienced user)
P: “What can you do?”	service	robot is switching to the verbosity level ‘novice user’
H: “I could offer some services such as fetching and carrying objects or simply entertain you!”	service	offering detailed information about capabilities; would have been extended if the next user input would not have been understood
P: “Take over glass!”	service	service task not yet completely specified
H: “Should I take over the glass with my left hand?”	confirmation	implicit feedback is given to the user by incorporating ‘glass’ in the question; context switching for better recognition
P: “No!”	confirmation	now the robot assumes that the right hand should be used
H: “I will take over the glass with my right hand. Please, hand over glass!”	service	bringing the gripper into a position and orientation that allows taking over a glass; if the user wishes to cancel he could say ‘no’ or ‘cancel command’
P: *doing nothing*	service	robot is waiting for user to act as previously instructed
H: “Please place the glass between my fingers and touch them!”	service	incremental prompt: more information is given to the user to ensure proper interaction through kinesthetic sensing
P: *tricking the robot by touching the fingers but not giving the glass*	service	the robot notices that the user did not hand over an object because the gripper did not fully close
H: “Please try again to hand over glass or say ‘cancel command’ ”	service	since something went wrong during the first trial other options are explicitly listed
P: *placing glass between fingers, thereby slightly touching them*	service	robot notices the touching of its fingers and grasps the glass; if no user interaction had happened the robot would have returned to its initial state
H: “What else can I do for you?”	service	continuing to offer services to a novice user (prompt would be omitted otherwise)
P: “Place it onto the table!”	service	‘it’ is replaced with ‘glass’ because the glass is the most recent object the user has referred to
H: “I have detected two tables. Please specify one!”	object attrib.	implicit feedback is given to the user by incorporating the number of tables detected in the utterance; context switching
P: “The small one!”	object attrib.	key word spotting for attributes such as left, right, small, ...
H: “I will place the glass onto the small table. Do you want me to do this task?”	confirmation	implicit feedback is given to the user by incorporating the attributes of the table in the utterance; confirmation is required before executing this time-consuming task
P: “Yes, please!”	confirmation	key word spotting for ‘Yes’ or ‘No’
H: “I am approaching the table.” (more utterances until task completion)	idle	since the verbosity level is still set to ‘novice user’, the robot is commenting its actions aloud, otherwise it would execute its task silently

Figure 4: Excerpt from a dialogue between a human and *HERMES* to take over an object and place it onto a table. In its course *HERMES* learns that it is interacting with a novice user who needs much feedback for interacting correctly. It should be noted how often contexts are switched, depending on the robot’s expectations. This makes the speech recognition much more dependable and robust, especially in the presence of ambient noise.

any emotions as researched by [Hara et al. 1998] with the realistic robot head MARK II and [Breazeal, Scassellati 1999] with the caricature of a human face KISMET. [Thrun et al. 2000] reported that caricatured facial pseudo-emotions proved to be one of the most appreciated aspects of their museum tour guide MINERVA. It is arguable whether robotic assistants should pretend to have emotions to facilitate (social) interaction, and space is insufficient to fully discuss this issue, but from a more practical perspective there are good reasons to implement at least some human-like behaviors to support communicative goals.

For example, to give appropriate way descriptions to a dialogue partner, *HERMES* calculates a path from its own current location to the goal location based on its topological map of the environment and deduces a way description. Since humans cannot measure distances or angles without special devices, it does not use such quantities in its description and starts with a deictic expression, such as "Go along this corridor". The great advantage of using deictic expressions with way descriptions is that they can be understood regardless of the position and orientation of the robot's dialogue partner and are, therefore, generally applicable. However, in general, they can only be interpreted relative to a (usually) extralinguistic context of the utterance, such as the gestures of the speaker. An example is depicted in Figure 1 showing *HERMES* pointing and looking to its left hand side supporting the utterance "Follow *this* corridor".

To attract people's attention, *HERMES* is able to wave with one or both of its arms while talking aloud; it may then answer questions, play music, or it may directly address a passer-by and initiate a hand shake. But more importantly from our viewpoint, *HERMES* uses its arms, body, head and eyes to support other communicative goals, such as supporting deictic expressions or exchanging objects with humans. Combined communicative, motor and sensor skills may also serve to compensate for its current inability to track a human face or hand visually.

Interactions, such as hand shaking and exchanging objects, require tactile sensor skills. In combination with motor skills, such as gross arm positioning, objects can be received from, or given to, people. 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. This arm positioning is accompanied by an utterance asking the human to do whatever is required to finish the interaction.

Kinesthetic sensing has proved to be a very reliable means of guiding the human and the robot through the interaction. By intelligently processing joint angle encoder values and motor currents the robot is able to detect touch events, i.e., tensions on the robot structure or torques at the joints that do not result from internal motion requests, but most probably from external circumstances. 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. Depending on the situation and on which joint is signaling, the touch event may indicate that, e.g., a human has touched the robot's finger or that a collision has occurred.

4 Real-World Experiments and Results

Multiple interaction experiments have proved the suitability of our approach and the long-term dependability of *HERMES'* communication system in the real world with multiple naive partners. The robot was presented at trade fairs, in television studios and at various demonstrations in our institute environment. The dialogue and associated human-robot interactions illustrated in Figures 4 and 5 may serve as an example how *HERMES* interacts with people. Contexts are frequently switched depending on the prevailing situation. Prompts implicitly incorporate a verification of the user's utterance, and the verbosity level is automatically adapted to the user's apparent knowledge and experience. Whenever a command is incomplete (missing command arguments) or ambiguous (too many arguments or imprecise description), a specific dialogue is initiated to resolve the problem. It is important to note that the robot is always in charge of the current dialogue (except in an emergency).

Direct interaction with people is greatly improved by *HERMES'* kinesthetic sense. It enables *HERMES* to safely manage object exchanges, object delivery and hand shaking with novice users. By wandering around and asking passers-by for the names of specific locations *HERMES* is able to build an attributed topological map of an unknown building. On this basis it is then able to give way descriptions that are accompanied by arm and head gestures to humans (Figure 1).

From 10/01-04/02 *HERMES* was field tested in the Heinz Nixdorf MuseumsForum (HNF), in Paderborn, Germany. In the special exhibition "Computer.Brain" HNF presented the current status of robotics and artificial intelligence and displayed some of the most interesting robots from international laboratories, including *HERMES*.

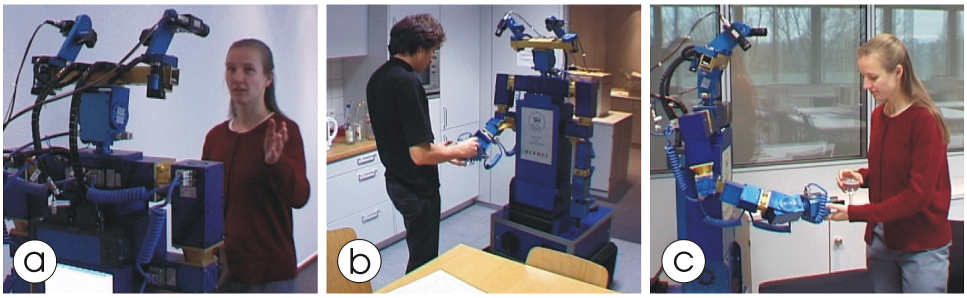

Figure 5: *HERMES* executing service tasks in the office environment of the Heinz Nixdorf MuseumsForum: (a) dialogue with an *a priori* unknown person, where *HERMES* accepts the command to get a glass of water from the kitchen and to carry it to the person's office; (b) asking a person in the kitchen to hand over a glass of water; (c) taking it to the person's office and handing it over

We used this opportunity of having *HERMES* in a different environment to carry out experiments involving all of its skills, such as vision-guided navigation and map building in a network of corridors; driving to objects and locations of interest; manipulating objects, exchanging them with humans or placing them on tables; kinesthetic and tactile sensing; and detecting, recognizing, tracking and fixating objects while actively controlling the exposure time of the cameras. Visitors and museum employees have been most impressed by *HERMES’* plug-and-run capabilities, its human-friendly intuitive interaction interface and its dependability. *HERMES* was able to chart the office area of the museum from scratch upon request and delivered services to *a priori* unknown humans (Figure 5). 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 visitors. *HERMES* chatted regularly with employees and international visitors in three languages (English, French and German). Topics covered in the conversations were the various characteristics of the robot (name, height, weight, age, etc.), 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, driving to a table and placing the glass onto it; playing the visitors’ favorite songs and telling jokes that were also retrieved from the Web.

5 Summary and Conclusions

By integrating various sensor modalities including vision, a haptic sense and hearing, a robot may be built that displays intelligence and cooperativeness in its behavior and communicates in a user-friendly way. This was demonstrated in experiments with a complex robot designed according to an anthropomorphic model. A special kind of behavior-based system architecture was implemented to control the robot. Its main idea is to activate and coordinate dynamically the robot’s skills and behaviors (including communication behaviors) based on an assessment of the situation being perceived by the robot at a particular moment. This concept proved to be effective, although it 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 micro-controllers and digital signal processors embedded in a single PC, in combination with the concept of skills for organizing and distributing the execution of behaviors efficiently among the processors, is able to meet these demands. We claim that the demonstrated robustness and dependability of the robot is largely due to this situation-oriented approach. Another consequence of the approach is that the robot accepts orders that would be given to a human in a similar way.

Human-robot communication is primarily based on speech that is recognized speaker-independently without any prior training of the speaker. A high degree of robustness is obtained due to the concept of situation-dependent invocations of grammar rules and word lists. Human-robot interaction is further facilitated by the robot’s kinesthetic sense, e.g., when exchanging objects with a human, and its ability to use gestures, e.g., for giving way instructions.

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

Acknowledgment

We thank Mr. Gheorghe-Costel Balica for programming major parts of *HERMES’* man-machine interface and for the many fruitful discussions we had.

References

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.

Breazeal, C.; Scassellati, B. (1999): How to build robots that make friends and influence people. IEEE/RSJ Int. Conf. on Intel. Robots and Systems (IROS-99), Kyongju, Korea, Vol.2, pp. 858-863.

Hara, F.; Kobayashi, H.; Iida, F.; Tabata, M. (1998): Personality characterization of animate face robot through interactive communication with human. In IARP 1st International Workshop in Humanoid and Human Friendly Robotics, Tsukuba, Japan, pp. 1-10.

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.

Lemon, O.; Bracy, A.; Gruenstein, A.; Peters, S. (2001): Information states in a multi-modal dialogue system for human-robot conversation. 5th Workshop on Formal Semantics and Pragmatics of Dialogue (Bi-Dialog 2001).

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

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

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