

Natürlichsprachliche Steuerung eines anthropomorphen Serviceroboters mit Gesichts- und Tastsinn

Natural Language-Based Control of an Anthropomorphic Service Robot with Visual and Haptic Senses

Rainer Bischoff

Universität der Bundeswehr München, Institut für Meßtechnik

Werner-Heisenberg-Weg 39, 85577 Neubiberg

Telefon: +49-89-6004-3589, Fax: +49-89-6004-3074

Rainer.Bischoff@unibw-muenchen.de

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

Kurzfassung

Im ständig wachsenden Dienstleistungsbereich sollen mit sogenannten Servicerobotern oder persönlichen Robotern neue Märkte erschlossen werden. Entscheidend für den Erfolg und die Akzeptanz dieser Systeme ist eine weitgehend intuitive und damit auch für Laien verständliche Mensch-Roboter-Interaktion. Um diesem Ziel näherzukommen, wurde der Serviceroboter *HERMES* nicht nur in der Anzahl und Anordnung seiner Freiheitsgrade dem Menschen nachempfunden, sondern auch mit den wichtigen Sinnen Sehen, Tasten und Hören ausgestattet. Darauf aufbauende Fertigkeiten zur visuellen und taktilen Wahrnehmung sowie zur natürlich-sprachlichen Kommunikation wurden unter einer einheitlichen situationsbewertenden verhaltensbasierten Systemarchitektur zusammengefaßt. Dadurch läßt sich der Roboter schnell in neue, primär für Menschen geschaffene und von Menschen bevölkerte Umgebungen einbringen und den dort lebenden bzw. arbeitenden Menschen dienstbar machen.

Abstract

In the constantly growing service sector so-called service and personal robots will allow to open up new markets. Key to the successful introduction and acceptance of these systems is a largely intuitive and also for non-experts comprehensible man-machine interaction. To advance research in this field, the service robot *HERMES* has been built according to an anthropomorphic model, not only with respect to the number and configuration of its degrees of freedom, but also to its basic senses such as vision, touch and hearing. Built upon these senses skills for visual and tactile perception and natural language-based communication have been defined and integrated under a unifying situation-oriented and behavior-based system architecture. It allows an easy and fast introduction of the robot into new, populated living or working environments where it is to perform its services.

1 Einleitung

Viele Wissenschaftler arbeiten weltweit daran, Roboter zu bauen, die eine Vielzahl verschiedener Sinne aufweisen, sich sicher fortbewegen und intuitiv bedient werden können. Obwohl der Gesichts- und Tastsinn sowie Sprachverstehen neben der Aktuatorik Hauptbestandteile eines benutzerfreundlichen Roboters sind, hat man sie bis heute eher unabhängig voneinander betrachtet, hauptsächlich, weil sie eigenständige, umfangreiche Forschungsgebiete darstellen. Deshalb gibt es bis heute nur eine geringe Anzahl an Robotern, die menschenähnliche Manipulationsmöglichkeiten besitzen, die drei wichtigen Sinne Sehen, Tasten und Hören integrieren und autonom und autark in wirklichen Umgebungen arbeiten können.

Um benutzerfreundliche persönliche Roboter entwickeln zu können, müssen drei Hauptprobleme gelöst werden: (1) Entwurf und Integration von Sensoren und Aktuatoren, die es einem Roboter ermöglichen, seine Umwelt ähnlich wie ein Mensch wahrzunehmen und in ihr ähnlich wie ein Mensch zu agieren; (2) Verwirklichung einer Steuerungsarchitektur, die unter Berücksichtigung einer Vielzahl unterschiedlicher Sensorinformationen und auf der Grundlage der begrenzten Bewegungsmöglichkeiten der Aktuatorik es dem Roboter ermöglicht, für einen Benutzer Handlungen zielgerichtet und sicher auszuführen; (3) Entwicklung von Verhaltensmustern zur intelligenten Kommunikation und kooperativen Interaktion mit dem Benutzer.

Unser Vorschlag zur Lösung dieser drei Probleme ist es, den Roboter nach menschlichem Vorbild (anthropomorph) aufzubauen, ihn mit universell verwendbaren Sensormodalitäten auszustatten (z.B. Sehen, Tasten, Hören) und auf dieser Grundlage menschenähnliche sensomotorische und kommunikative Fertigkeiten und Verhaltensmuster zu implementieren. So wird es dem Roboter ermöglicht, mit Menschen auf intelligente Art und Weise zu interagieren und Objekte in seiner Umgebung zu manipulieren.

2 Vorschläge zur Lösung der Hauptprobleme bei Servicerobotern

2.1 Entwurf eines Roboters für Alltagsumgebungen

Zur Lösung des ersten Problems, der Konstruktion des Roboters, scheint es unter den gegebenen Randbedingungen plausibel zu sein, den menschlichen Körperbau und die menschliche Sensorik als Vorbild zu nehmen. Der Hauptgrund für eine solche Herangehensweise ist, daß ein persönlicher Roboter oder Serviceroboter seine Aufgaben dort erledigen soll, wo normalerweise Menschen leben und arbeiten, z.B. in Wohnungen, Büros, Laboren, Restaurants oder Krankenhäusern. Grundidee ist hier, nicht die Umgebungen an die Erfordernisse des Roboters anzupassen, sondern den Roboter an die vorgesehene Art von Umgebungen anzupassen. Diese Umgebungen sind auf die Eigenschaften und Bedürfnisse des Menschen optimiert, d.h. ergonomisch gestaltet worden, z.B. nach dem Platzbedarf eines Menschen, nach seiner Arbeits- und Blickhöhe und seiner Kraft und der Anzahl der Freiheitsgrade, die er zur Handhabung von Objekten einsetzt [Bischoff 1997].

2.2 Steuerung eines anthropomorphen Serviceroboters

Zur Lösung des zweiten Problems, der Steuerung eines Roboters mit mehreren, teilweise redundanten Freiheitsgraden in Sensorik und Aktuatorik, schlagen wir die Verwendung einer verhaltensbasierten Systemarchitektur vor. Verhaltensbasierte Systemarchitekturen werden mittlerweile allgemein als effiziente Grundlage autonomer mobiler Roboter angesehen [Arkin

1998]. Sie ermöglichen es, daß ein Roboter ein bestimmtes Ziel durch Verknüpfung mehrerer Verhaltensmuster erreicht, die er aus seinem Repertoire an Verhaltensmustern auswählt.

Das Hauptproblem beim Entwurf verhaltensbasierter Roboter ist die Frage, auf welcher Grundlage in jedem Moment das richtige Verhalten ausgewählt werden soll. Wir schlagen vor, für diese Entscheidung eine Vielzahl von Faktoren zu betrachten, die wir unter dem Begriff „Situation“ zusammenfassen. Das Konzept der „Situation“ umfaßt hier nicht nur die Objekte in der Umgebung und deren Bewegungszustände, sondern auch Zielvorgaben durch den Benutzer, roboterinterne Ziele wie z.B. Unfallfreiheit sowie die eventuell eingeschränkten Handlungsmöglichkeiten des Roboters [Graefe, Bischoff 1997].

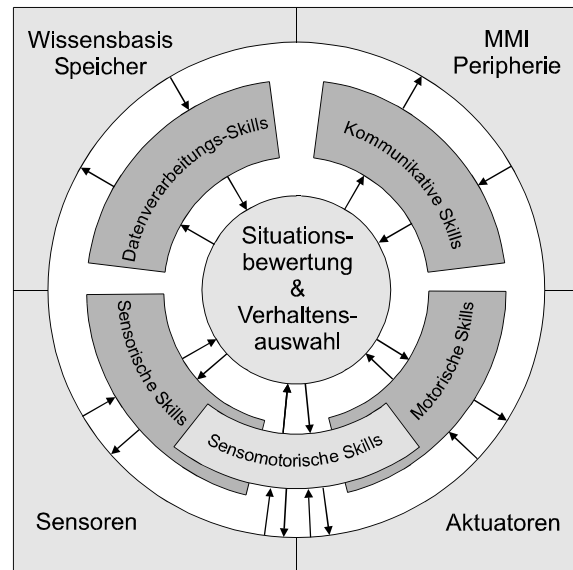


Abbildung 1: Systemarchitektur eines Service-roboters basierend auf den Konzepten Situation, Verhalten und Fertigkeiten („Skills“).

Damit die von uns entwickelte Systemarchitektur geeignet ist, mehrere Sensormodalitäten, Aktuatoren, Wissensbasen und eine benutzerfreundliche Mensch-Maschine-Schnittstelle unter einer einheitlichen Oberfläche zu vereinen, wurde ergänzend zum Situationskonzept das Konzept der Fertigkeiten („Skills“) eingeführt (Abbildung 1). Diese Fertigkeiten verbinden auf bidirektionale Weise das Situationsmodul („Situationsbewertung und Verhaltensauswahl“) als Kernstück des Systems mit allen Hardwarekomponenten (Sensoren, Aktuatoren, Speicher für die Wissensbasis und Peripheriegeräten zur Mensch-Maschine-Kommunikation). Da sie direkten Zugriff auf die Schnittstellen dieser Hardwarekomponenten haben, können sie das Situationsmodul mit bestimmten Informationen versorgen (z.B. mit Sensorwerten oder Objektkoordinaten), im Auftrag des Situationsmoduls bestimmte Ausgaben erzeugen (z.B. die Arme bewegen oder sprechen) oder auch bestimmtes Wissen verarbeiten bzw. abrufen (z.B. einen Pfad basierend auf Kartenwissen planen). Fertigkeiten können grob unterteilt werden in motorische, sensorische und sensomotorische Fertigkeiten (als Kombination der ersten beiden Fertigkeiten) sowie kommunikative Fertigkeiten und Datenverarbeitungsfertigkeiten.

Das Situationsmodul sammelt die von den einzelnen Fertigkeiten mitgeteilten Daten und Informationen über den Systemzustand und wird so in die Lage versetzt, Situationen zu bewerten und bestimmte Fertigkeiten gezielt auszuwählen und zu parametrieren. Darüber hinaus stellt es kognitive Fertigkeiten in Form des gesamten Systemmanagements zur Verfügung, d.h., es ist verantwortlich für die Planung einer bestimmten Abfolge von Verhaltensmustern und für die zum Ausführen der Verhaltensmuster notwendige Aktivierung bzw. Deaktivierung von Fertigkeiten. Das nach außen sichtbare Roboterverhalten wird also durch das bewußte Verknüpfen bzw. Aneinanderreihen elementarer Fertigkeiten erreicht.

2.3 Kommunikation und Interaktion mit Servicerobotern

Eine benutzerfreundliche Schnittstelle ist eine Grundvoraussetzung für Roboter, die den Menschen bei verschiedenen Tätigkeiten im Alltag unterstützen sollen. Zwei Probleme müssen

dabei gelöst werden: Erstens müssen sich Mensch und Roboter über einen geeigneten Kommunikationskanal verständigen, und zweitens müssen sie sich über den so ausgewählten Kommunikationskanal über gemeinsame Referenzpunkte in der Umgebung einigen.

Da natürliche Sprache der nächstliegende und natürlichste Kommunikationskanal für einen Menschen ist, liegt es nahe, Spracherkennung und -ausgabe in die meisten Serviceroboter zu integrieren. Sprache kann dazu verwendet werden, dem Roboter neue Aufgaben und Ziele vorzugeben und die Verhaltensausführung unmittelbar zu modifizieren. Damit die Roboter jedoch als kooperative und mitdenkende Partner akzeptiert werden, müssen sie nicht nur eindeutige und vollständige Befehle verstehen, sondern auch Mehrdeutigkeiten auflösen und unvollständige Sätze, die sich aus einer Dialogsituation heraus ergeben können, ergänzen können. Um diese Probleme zu lösen, sollte der Roboter zum einen ein Verständnis seiner augenblicklichen Situation haben und zum anderen fehlende Informationen vom Benutzer gezielt erfragen können.

Eine weitere wichtige zu klärende Frage ist, wie Menschen und Roboter ihre Kommunikation und Interaktion auf eine gemeinsame Grundlage stellen können. Je stärker sich die sensorischen Fähigkeiten des Roboters von denen des Menschen unterscheiden, desto schwieriger werden sich die Kommunikation und die Interaktion zwischen dem Menschen und dem Roboter gestalten. Größtmögliche Menschenähnlichkeit des Roboters bezüglich seiner sensorischen Fähigkeiten ist deshalb erwünscht, aber bisher nur eingeschränkt realisierbar.

Trotzdem können Roboter und Mensch miteinander kommunizieren, wenn sie sich auf gemeinsame Referenzpunkte in der Umgebung und eine gemeinsame Namensgebung einigen. Beispielsweise könnte Sprache dazu verwendet werden, Orts- bzw. Objektnamen als Grundlage für eine Zusammenarbeit zu verwenden. Diese Übereinkunft verlangt vom Menschen kein Verständnis für die vom Roboter verwendete Repräsentation der Welt, solange sich der Roboter die vereinbarten Bezugspunkte in seiner Repräsentation merken und wenn erforderlich darauf zurückgreifen kann.

Der hier vorgeschlagene Ansatz, der auf einem anthropomorphen Roboter mit der Fähigkeit, sein Verhalten situationsabhängig zu steuern, beruht, ist geeignet, dieses Kommunikations- und Interaktionsproblem zu lösen. Da Mensch und Roboter in derselben Umwelt agieren und diese mit vergleichbarer Sensorik wahrnehmen, werden die von beiden wahrgenommenen Situationen ähnlich sein. In diesem Fall wird der Roboter in der Lage sein, Befehle entgegenzunehmen, die in ihrer Form und ihrem Abstraktionsniveau vergleichbar mit den Befehlen sind, die man einem Menschen in vergleichbarer Situation geben könnte. Der Roboter würde dann als intelligent und kooperativ angesehen werden, und man könnte mit ihm sehr leicht kommunizieren.

3 Realisierung der Lösungsvorschläge

Um die vorgestellten Konzepte und Lösungsvorschläge in der Praxis untersuchen und bewerten zu können, haben wir den anthropomorphen Roboter *HERMES* mit 22 Freiheitsgraden entworfen und konstruiert. Aus technischen und wirtschaftlichen Gründen hat er keine Beine, sondern ein omnidirektionales Fahrwerk, das ihm große Manövrierfähigkeit auf ebenem Boden verleiht. *HERMES* hat die Größe (1,80 m) und Gestalt eines erwachsenen Menschen. Sein Manipulationssystem besteht aus einem abknickbaren Oberkörper mit zwei Armen und Zweifin-

gergreifern. Seine beiden „Augen“ (monochrome Videokameras) sind unabhängig voneinander und hochdynamisch in Azimut und Elevation steuerbar und auf einem gemeinsamen Träger montiert, der ebenfalls geschwenkt und geneigt werden kann. *HERMES* Propriozeptoren umfassen unter anderem verschiedene Winkelkodierer, Stromwandler und Temperatursensoren. Ein kinästhetischer Tastsinn ist durch den fortlaufenden Vergleich von aktuellen Propriozeptorsignalen mit gewünschten oder in der betreffenden Situation erwarteten Signalen realisiert [Bischoff, Graefe 1999]. Die Stoßfänger, welche die gesamte Fahrplattform bis zu einer Höhe von 30 cm umgeben, sind auf Basis leitfähiger Schaumstoffe realisiert worden. Sie schützen Menschen und Material bei möglichen Kollisionen und lassen darüber hinaus Rückschlüsse auf die Art und den Ort der Kollision sowie die dabei ausgeübte Kraft zu.

Ein hierarchisch organisiertes Mehrprozessorsystem bestehend aus 25 Mikrocontrollern, vier digitalen Signalprozessoren und einer Pentium CPU ist für die echtzeitfähige Informationsverarbeitung verantwortlich [Bischoff 1997]. Ein „Roboter-Betriebssystem“ ist entwickelt worden, welches das Versenden und Empfangen von Nachrichten über verschiedene Kanäle zwischen den unterschiedlichen Prozessoren regelt. Alle Prozesse laufen prinzipiell asynchron, können aber über Ereignisse und Nachrichten synchronisiert werden.

Die Gesamtsteuerung wird durch einen endlichen Automaten realisiert, der auf priorisierte Interrupts und Nachrichten reagieren kann. Nach dem Einschalten befindet sich der Roboter im Zustand „Warten auf die nächste Missionsbeschreibung“. Eine Missionsbeschreibung kann entweder über E-Mail, direkt über Tastatur oder über Sprache übermittelt werden. Sie besteht aus konkret benannten und parametrisierten elementaren Fertigkeiten oder aus bestimmten natürlichsprachlichen Befehlen, deren Interpretation im jeweiligen Kontext betrachtet komplexes Roboterverhalten hervorruft.

Die Mensch-Maschine-Schnittstelle, dessen Kernstück ein kommerziell erhältliches Sprachein- und -ausgabesystem ist, wurde auf einem PC unter Windows NT realisiert. Mit Hilfe von vordefinierten Grammatiken und Wortschätzen wird kontinuierlich gesprochene Sprache sprecherunabhängig erkannt, so daß unterschiedliche Benutzer ohne besonderes Training mit dem Roboter kommunizieren können. Durch kontextsensitives Umschalten zwischen verschiedenen auf bestimmte Situationen angepaßte Grammatiken und Wortschätze wird eine große Robustheit bei der Spracherkennung erreicht [Bischoff, Jain 1999].

4 Experimente und Ergebnisse

Ein Beispiel soll das Potential des vorgeschlagenen Konzeptes verdeutlichen, aber auch die Beschränkungen der derzeitigen Implementierung aufzeigen. Abbildung 2 zeigt exemplarisch, wie ein einfacher Hol- und Bringdienst aussieht. Der dazugehörige Dialog zwischen Roboter und Benutzer ist auf Englisch wiedergegeben, da zur Zeit lediglich ein englisches Sprachein- und -ausgabemodul in die Bedienoberfläche integriert ist.

Mensch: “Go to the secretariat!”

HERMES Situation ändert sich, sobald er einen Befehl empfängt. In diesem Fall wird ihm der Auftrag per E-Mail mitgeteilt. Nach dem Erhalt und der Interpretation der Nachricht beginnt *HERMES*, einen Weg von seiner Basisstation zum Sekretariat zu planen. Der Weg wird von einer Datenverarbeitungsfertigkeit in Form von Verhaltensmustern geplant (z.B.: Verlasse Labor, biege rechts ab, biege an der zweiten Kreuzung links ab, biege rechts ab, halte an der

zweiten Tür auf der rechten Seite an). Für die Ausführung der Verhaltensmuster werden keine koordinatenbasierten Fahrbefehle benötigt. Wenn während der Navigation nichts Unvorhergesehenes passiert (z.B. Weg blockiert), erreicht der Roboter sein Ziel vor der Tür zum Sekretariat, wo er auf sich aufmerksam macht.

HERMES: “Hello! What can I do for you?” – Mensch: “Take over tray!”

Aus seiner Objekt-Datenbasis weiß der Roboter, wie ein Tablett aussieht und wie er es greifen muß, damit Objekte nicht herunterfallen können. Da der Roboter noch nicht die Fertigkeit besitzt, die augenblickliche Pose des Tabletts zu erkennen, bringt er seinen Arm einschließlich des Greifers in eine Konfiguration, die einem Menschen eine leichte Übergabe ermöglicht.

HERMES: “Please hand over tray!”

Jetzt wartet *HERMES* auf eine mit Hilfe seines kinästhetischen Tastsinns erkennbare Berührung, die ihm signalisiert, die Greiferbacken zu schließen, um so das Tablett zu greifen. Nach dem erfolgreichen Greifvorgang fordert er den Menschen zu weiteren Instruktionen auf.

HERMES: “What else can I do for you?”

Mensch: “Place it onto the table in the kitchen, then go home.”

Um diese komplexe Instruktion zu verstehen, muß sie zunächst in ihre elementaren Bestandteile zerlegt und auf Vollständigkeit überprüft werden. Zuerst muß *HERMES* herausfinden, auf welches Objekt sich das Pronomen „it“ bezieht, anschließend muß er überprüfen, ob er das Objekt („table“), auf dem etwas abgelegt werden soll, und den Ort („kitchen“), an dem er dieses Objekt finden soll, kennt. Da der Roboter mit der grundlegenden Fertigkeit ausgestattet ist, Tische zu erkennen, und in diesem Fall nur der Ort unbekannt ist, bittet der Roboter um die Vervollständigung des fehlenden Wissens.

HERMES: “I do not know the place ‘kitchen’. Please, show me the way!”

Mensch: “Follow the corridor to your right, turn left at the first intersection, turn right. The kitchen will be located at the end of the corridor.”

Anhand der Wegbeschreibung und seines augenblicklichen Kartenwissens kann der Roboter verifizieren, ob er den Ort bereits unter einem anderen Namen kennt und kann in diesem Fall die Liste der Namen für diesen Ort ergänzen. Kennt er den Ort nicht, wird er sich anhand der Wegbeschreibung auf die Suche nach diesem Ort machen und sein Kartenwissen entsprechend erweitern. So kann sich der Roboter im Laufe der Zeit alle auftragsrelevanten Orte unter

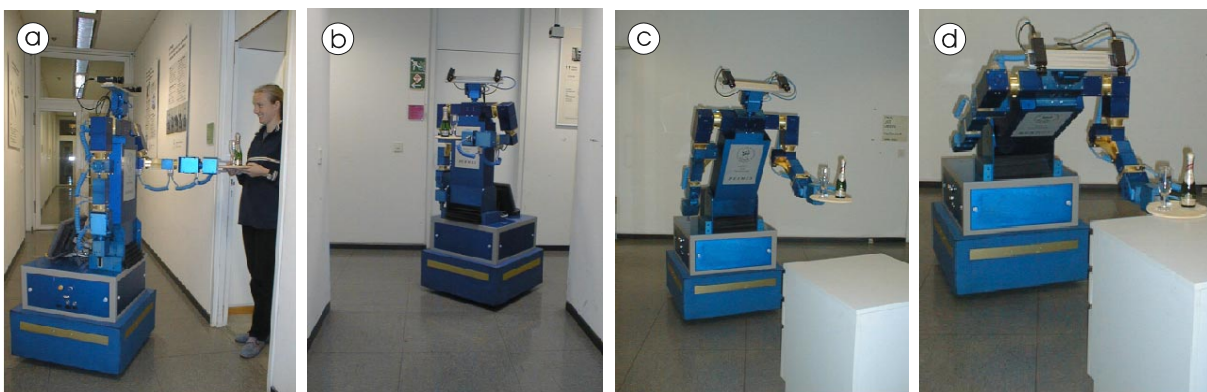


Abbildung 2: *HERMES* erledigt einen Dienstleistungsauftrag: a) Kommunikation und Interaktion mit einem Nutzer; Übernahme eines Tabletts; b) Navigation im Wegenetz des Gebäudes zum befohlenen Zielobjekt c) Annäherung an einen Tisch (Zielobjekt) d) Absetzen des Tabletts auf dem Tisch

den für die jeweiligen Benutzer maßgeblichen Namen merken. Seine interne Repräsentation spielt für den Benutzer dabei keine Rolle. Wenn alle Befehlsparameter bestimmt worden sind, führt *HERMES* den in elementare Kommandos zerlegten Auftrag aus. Zunächst fährt er in die Küche, sucht dort nach einem Tisch, fährt mit Hilfe des in [Bischoff, Graefe 1998] beschriebenen sichtbasierten Andock-Manövers an den Tisch heran und setzt das Tablett mit Hilfe seines Tastsinns und einiger vorprogrammierter Bewegungen (motorischen Fertigkeiten) ab. Anschließend fährt er zurück zu seiner Basisstation und wartet auf neue Instruktionen.

5 Zusammenfassung

Der von uns realisierte Serviceroboter *HERMES* besitzt bereits mehrere motorische, sensorische, sensomotorische, kommunikative und kognitive Fertigkeiten, die menschenähnliches Verhalten in bestimmten Situationen generieren können. Dazu gehören u.a. die sichtbasierte Navigation in einem Netzwerk von Korridoren, das Erkennen von Andockobjekten und die Annäherung an solche Objekte, die Übernahme von Gegenständen von einem Menschen sowie deren Übergabe an einen Menschen und das Abstellen von Gegenständen auf Tischen. Über natürlichsprachliche Befehle kann mit dem Roboter kommuniziert werden, beispielsweise um einfache Hol- und Bringdienste aufzutragen. Dies geschieht entweder in Form von geschriebenem Text, z.B. über E-Mail, oder mittels gesprochener Sprache. Das dialogbasierte überwachte Lernen einer Einsatzumgebung zum Aufbau gemeinsamer Referenzpunkte in der Umgebung zwischen mehreren Benutzern und dem Roboter ist ebenfalls möglich.

Die ersten Experimente und einige öffentliche Präsentationen haben gezeigt, daß die menschenähnliche Gestalt des Roboters und seine animierende Sprachausgabe die Menschen zu natürlicher Interaktion und Kommunikation ermuntern. Selbst Laien konnten problemlos mit dem Roboter natürlichsprachlich kommunizieren und ihm Aufträge erteilen. Dies untermauert unsere Annahme, daß zukünftige Roboter umso menschenähnlicher gestaltet und ihre Steuerungen umso bedienungsfreundlicher realisiert werden müssen, je näher der Kontakt mit Menschen ist. Dazu benötigt der Roboter neben einer menschenähnlichen Anordnung und Anzahl an Freiheitsgraden jedoch auch menschenähnliche sensorische Fähigkeiten, damit Mensch und Roboter die Situationen, in denen sie sich befinden, auf einem vergleichbaren Abstraktionsniveau wahrnehmen können. Ist diese Grundvoraussetzung erfüllt, können auch Laien mit dem Roboter intuitiv und effizient kommunizieren und kooperieren.

6 Literatur

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

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

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

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

Bischoff, R.; Jain, T. (1999). Natural Communication and Interaction with Humanoid Robots. Second International Symposium on Humanoid Robots. Tokyo, Japan, October 1999, pp. 121-128.

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