Human-Robot Collaboration: a Survey.

Article *in* International Journal of Humanoid Robotics · March 2008

DOI: 10.1142/S0219843608001303 · Source: DBLP

Dirk Wollherr
Technische Universität München

Martin Buss
Technische Universität München

International Journal of Humanoid Robotics

Human-Robot Collaboration: A Survey

Andrea Bauer, Dirk Wollherr, Martin Buss
Institute of Automatic Control Engineering (LSR)
Technische Universität München
80290 Munich
Germany
ab@tum.de; dw@tum.de; mb@tum.de

Received 01.05.2007
Revised 29.09.2007
Accepted Day Month Year

As robots are gradually leaving highly structured factory environments and moving into human populated environments, they need to possess more complex cognitive abilities. They do not only have to operate efficiently and safely in natural, populated environments, but also be able to achieve higher levels of cooperation and communication with humans. Human-robot collaboration (HRC) is a research field with a wide range of applications, future scenarios, and potentially a high economic impact. HRC is an interdisciplinary research area comprising classical robotics, cognitive sciences, and psychology. This article gives a survey of the state of the art of human-robot collaboration. Established methods for intention estimation, action planning, joint action, and machine learning are presented together with existing guidelines to hardware design. This article is meant to provide the reader with a good overview of technologies and methods for HRC.

Keywords: Human-robot collaboration; intention estimation; action planning; machine learning.

1. Introduction

Human-robot Collaboration (HRC) is a wide research field with a high economic impact. Robots have already started moving out of laboratory and manufacturing environments into more complex human working environments such as homes, offices, hospitals and even outer space. HRC is already used in elderly care[1], space applications[2], and rescue robotics[3]. The design of robot behaviour, appearance, cognitive, and social skills is highly challenging, and requires interdisciplinary cooperation between classical robotics, cognitive sciences, and psychology. Humans as nondeterministic factors make cognitive sciences and artificial intelligence important research fields in HRC.

This article refers to human-robot collaboration as opposed to human-robot interaction (HRI) as these two terms hold different meanings[4]. Interaction is a more general term, including collaboration. Interaction determines action on someone else. It is any kind of action that involves another human being or robot, who does not necessarily profit from it. For an overview of HRI see Fong et al.[5] and Kiesler et al.[6]. Yanco et al.[7] provide a taxonomy to classifying HRI by eleven categories, including task type, robot morphology, interaction roles, time, and space. Collaboration means working with someone on something. It aims at reaching a common goal.

Humans and robots collaborating on a common task form a team. A team is defined[8] as a small number of partners with complementary skills who are committed to a common purpose, performance goal, and approach for which they hold themselves mutually accountable. The same holds for human-robot teams where the partners are humans and robots, committed to reach a common goal through collaboration.

Efficient collaboration requires a common plan for all involved partners. To gain a joint intention they need to know the intentions of the other team members and what they are doing. Based on that knowledge a robot can plan its own actions that will eventually lead to fulfil the joint intention and reach a common goal. Therefore they need the abilities of perceiving and comprehending their environment, decision making, planning, learning, and reflection, which will be used as a definition of cognition in this article

Fig. 1 gives an overview of the internal mechanisms of a cognitive robot leading to joint actions and collaboration. This is an example of an architecture for collaboration. Other architectures may have different sequences, but the main mechanisms are basically the same. The environment and the partners are observed by sensors. This sensor data is processed to gain an understanding of the environment and provides perception. The perceived data is used firstly to learn and expand the own knowledge, then to gain an understanding of the state of the environment and the partners, and to estimate the intention of the partners. When partners are collaborating, a joint intention is retrieved from the single intentions. A set of actions leading to fulfil the joint intention is found by action planning. At last actions are taken either by single partners or jointly that lead to transitions in the state of the environment. The loop is closed, as the robot observes the actions of itself and of others and the change in the environment. The methods indicated in grey in Fig. 1 are further analyzed in the following chapters. Sensors and perception have to be chosen and implemented individually depending on the purpose and desired abilities of the robot. The knowledge database has to be programmed and trained, according to the desired abilities. Perception and knowledge database will not be further discussed, given the various possibilities of those methods and the limited space in this article.

This article gives an overview of HRC. Components and methods that robots need for HRC are analyzed. The article gives only brief abstracts of the single methods, due to the limited space, though each of those topics could fill separate survey articles. In Section 2 mechanisms leading to collaboration are described, consisting of intention estimation and joint intention in Section 2.1, action planning

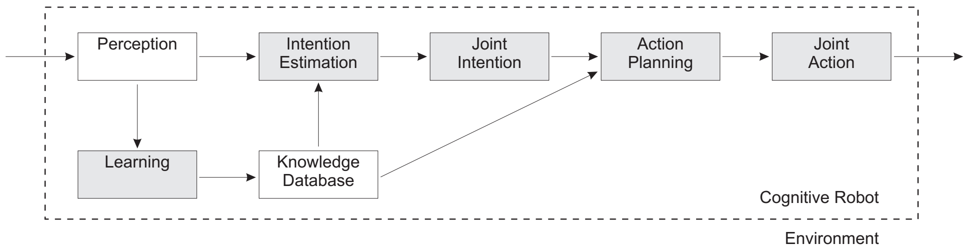

Fig. 1. Overview of the mechanisms leading to joint action.

in Section 2.2 and joint action in Section 2.3. Section 2.4 gives an overview of machine learning methods, that are used in the various fields of joint action. Finally in Section 3 thoughts about robot hardware design for HRC are given. In Section 4 examples of applications for HRC are presented.

2. Collaboration

HRC can be realized through actions of the individual partners aiding in fulfilling a joint intention or joint actions carried out jointly by multiple partners. Steps leading to joint action are joint intention, action planning and action. To work cooperatively on something the partners need to agree on a common goal and a joint intention to reach that goal. In human-robot collaboration it will be usually the human who states that goal, while it is the task of the robot to assist the human and take on the humans intention as its own and therefore as the joint intention of all. The human displays an intention to reach a certain goal. It is the robots task to estimate this intention which may be communicated by the human explicitly or implicitly. The intention of the robot becomes to help the human reach his or her intended goal. Thus human and robot obtain a joint intention to reach a certain goal.

When two or more partners have a joint intention they have to plan their actions according to the partners actions. These actions combine to joint actions. To plan an action the peer needs information about the actions and intention of the peers as well as knowledge about the abilities of all partners and about the state of the environment. Based on that knowledge a proper action has to be chosen out of a set of possible actions, that aids in fulfilling the joint plan and not obstructing any peer in its own sub plan. Subsequently that action can be executed either by one partner alone or jointly with other partners. If one of the partners acts in a way that was not foreseen by the other peers, actions have to be re-planned. Accordingly intention needs to be re-estimated all the time as states and intentions change over the time.

2.1. *Intention estimation and joint intention*

A team is considered a collection of partners to which the mental state of joint intention can be assigned. Joint intention is defined as a joint commitment to perform a collective action while in a certain shared mental state[9]. To obtain a joint intention the partners have to be able to estimate intentions of others and agree on reaching the intended goal jointly. In a human-robot team it will mostly be the human who assigns a goal for the team and who therefore has an intention to reach that goal. In order to assist the human in achieving a set goal, it is the task of the robot to estimate this intention and to act accordingly. A person can communicate his or her intention either deliberately by explicit communication or implicitly by actions.

In Fig. 2 ways of communicating intention are shown. Some of these intentions are communicated explicitly and some are communicated implicitly and sometimes unconsciously, indicated in grey.

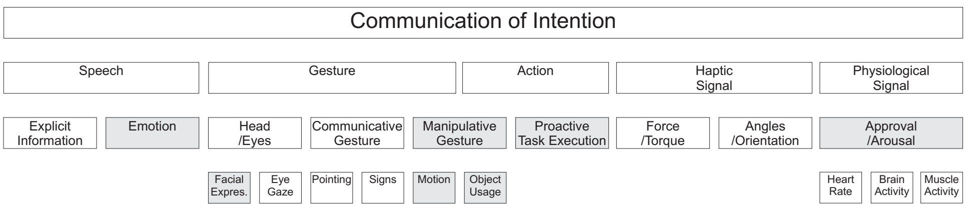

Fig. 2. Main ways of communicating intentions. Implicit communication is marked in grey.

The main ways of communicating one's intention are speech, gesture, actions, haptic signals and physiological signals. Speech provides explicit information about intentions through words and sentences, but can also provide information about emotions implicitly. Gestures are another means of communication. There are head, arm and body gestures. Head gestures are mostly facial expressions that often occur involuntarily and eye gaze direction that hints to what object a partner is referring to. Explicit hand and body gestures are communicative gestures that can be pointing gestures, primitive signs, or sign language gestures, holding complex information. Implicit hand and body gestures are called manipulative gestures. Manipulative gestures are actions partners take in the environment, these can be manipulating objects or motions. Though a partner does not necessarily want to communicate through these manipulations or motions, intention can be derived by others nevertheless. Another action that can convey intention is proactive task execution, which is used by one partner to detect the exact intention of another partner through provoking an action. When partners are connected haptically either directly or through an object they both manipulate, intentions can be communicated haptically. Haptic communication can occur through applied forces and torques or joint angles and orientations. Intentions are also communicated involuntarily through physiological signals such as heart rate, brain or muscle activity, they can give information about human approval and arousal. All of these ways of communicating intentions are described below more precisely.

2.1.1. *Explicit communication*

Using explicit communication a partner makes sure the peer receives certain information about his or her intention. Explicit communication may happen through diverse channels, such as speech, gesture, or haptics.

To humans speech is a very natural way of communication and it can convey a huge amount of complex information. However it takes a big effort for a computer or robot to extract that information. An acoustic signal, captured by a microphone has to be converted to a set of words through **speech recognition**. The theoretical bases for speech recognition have been established by Davis[10] and Fry[11]. For speech recognition the captured acoustic data is mapped to frequency domain via a Fourier transform then the most likely sequence of phonemes and subsequently the most likely words composed of those phonemes are found e.g. by hidden Markov models.

If more complex information is needed the recognized sequences of words have to be analyzed further, **speech comprehension** needs to be done. The most important approaches to speech comprehension are semantic networks[12] and conceptual dependency[13]. A detailed overview of the history of spoken language communication with computers is given by Pieraccini[14]. Furthermore emotion of the speaker can be detected in human speech[15].

Not only can robots understand human speech but they can also **synthesize speech** to convey information and intention themselves. An overview of speech synthesis and speech recognition is given by Nugues[16]. It is even possible for a robot to express emotion though speech[17] and thereby convey intention implicitly.

Methods for speech recognition and speech synthesis are advanced and even available as commercial products. Whereas the following methods of intention estimation are not jet technically mature and will need more developing until robustly deployable.

Another very basic manner of communicating intention is by producing visual signals such as gestures and facial expressions. Robots can derive intentions from gestures through **gesture recognition**. Gestures are classified into manipulative and communicative gestures, the communicative ones are subdivided into pointing gestures, primitive signs and sign languages. Manipulative gestures[18] are often done unconsciously, and implicit communication of intention happens through performed actions. Communicative gestures are meant to tell the peer explicitly about the own intention, though even for humans it may sometimes be hard to interpret these. Pointing gestures are supposed to call the peer's attention to a certain object or location. Whereas some gestures are primitive signs that have a predefined meaning. Examples for primitive signs are stop, start, and turn. A widely used form of gestures are all sign languages used by deaf people in lieu of speech. Sign languages are the most developed and numerous gestures with well defined rules and grammar. Gestures are created by static hand or body pose, or by physical motions. Gestures can be recognized and translated into either symbolic commands or trajectory motion commands by robots. Input modalities for gesture data are touch screens, data-gloves or camera systems, the last being the most natural and unobtrusive to humans.

Two methods for visual interpretation of hand gestures are reviewed by Pavlovic[19]. The first method uses a 3D hand and/or arm model to determine gestures through joint angles and palm position. The other approach is appearance based using parameters such as images, image geometry, image motion, fingertip position, and motion to identify the observed gesture. A System developed by Murakami inputs the gesture data via a data-glove[20]. The gestures are identified by a neural network which has been trained on a finger alphabet containing 42 symbols. The ALIVE II system[21] identifies full body gestures, through basic image processing techniques. The system uses kinematic information from a gesture as part of a control command. Some HRI approaches use predefined gestures, which the robot may be able to recognize and choose an appropriate action[22]. One of the most structured sets of gestures are those belonging to any of the several sign languages. In sign language, each gesture already has assigned meaning , and strong rules of context and grammar are applied. To identify American sign language gestures Starner and Pentland's system uses an Hidden Markov Model method[23]. The system is able to recognize forty sign language gestures by running acquired hand feature vectors through all possible sets of five-word sentences.

Important information about the human's intention can also be derived from head gestures such as eye gaze direction[24] or facial expressions[25]. Humans often express their emotions through facial expressions. By interpreting these expressions robots can gain information about approval or mood. From eye gaze direction robots can derive what the human is focussing on. Some robots also are able to convey their intentionality through facial expression[26,27] and thereby act socially. Also some robots are able to generate hand and arm gestures[28] to communicate with their partners.

Haptic communication is another way of conveying intention in HRC. For example two partners carrying a big object together may know the peer's intended direction of motion from the applied forces or torques, this is researched by Kosuge[29,30]. Both the human and the robot can act as the leading part, the other partner estimates the peer's intention and by acting to fulfil this intention becomes the follower. Besides force and torque, angles and orientation can be used as information in haptic communication. Humans and robots can not only collaborate haptically by manipulating objects but as well interact directly for example in dancing[31,32] or handshaking[33].

2.1.2. *Implicit communication*

In implicit communication of intention there is a lot of interpretation to be done by the peers. One partner may derive the intention of others from watching their actions and behaviour, that is **manipulative gestures**. This is a very complex and difficult task and requires a high level of cognition. Also if actions are to be recognized by visual sensors this entails a very advanced standard of image processing. Many researchers use vision techniques to recognize actions[34,35] as these are very natural and unobtrusive. Some actions and thereby the intention can be recognized by the objects that are being manipulated[36]. Humans are able to discern intentions of other humans by observing them in motion[37]. In particular a human judges another's action as arising from an intention only if they believe the other human desires a goal and that he or she can achieve it. Thus it seems very reasonable that a robot needs some model of the human partner to be able to tell whether he or she is able to achieve a desired goal and derive an intention.

Bobick[38] presents methods to discern simple movements, activities and complex actions. Movements are defined as motions whose execution is consistent and easily characterized by a definite space-time trajectory. The appearance of the motion can be described reliably in terms of the motion of the pixels in images. The pixel-based description of the motion under different viewing conditions is the only knowledge required to see the movement. An activity is described as a statistical sequence of movements, the recognition of which requires knowledge about both the appearance of each constituent movement and the statistical properties of the temporal sequence. Finally actions are defined to include semantic primitives relating to the context of the motion. For a system to recognize actions it must include a rich knowledge base about the domain and be able to hypothesize and evaluate possible semantic descriptions of the observed motion.

It is also possible to mark objects with RFID-tags and have the human partner wear a glove with an integrated RFID-reader to recognize, which objects are being used and hence derive the executed action[39]. However in a long term in a more natural working environment it is not practicable to mark every single object with an RFID-tag and furthermore the human is probably obstructed by the glove.

A way of communicating one's intention implicitly is to provoke a reaction of the peer by **proactive task execution**[40]. This can be utilized by both robots and humans to figure out which of multiple possible intentions the peer has. For example one partner holding an empty glass might want to put it down somewhere or want to get it refilled. To find out which of those possibilities applies in that moment, an other partner might approach the glass with a bottle of water. If the glass is withdrawn, the other partner most likely does not want to get it refilled.

Physiological signals can be measured to deduce the level of approval and arousal[41] and from that derive an intention. Suitable physiological signals are brain or muscle activity, heart rate, and galvanic skin response. Heart rate analysis[42] and multiple physiological signals can be used to estimate human stress levels and act accordingly. Approval and arousal can be derived from these and other unconscious signals that give away emotional states. However, measurements of physiological signals require devices attached to the human that may hinder natural movement.

Ultimately as the term intention estimation implies, the intention of the peer is not exactly measured but the most probable intention is found. Of course errors such as non-understanding or misunderstanding can occur. Non-understanding denotes the case when no information whatsoever can be derived. In the case of misunderstanding an intention is estimated but it is inconsistent with the real intention. Also an intention can change during interaction. This has to be kept in mind as intention estimation is crucial for joint intention which is the base of the entire HRC.

When intention is estimated correctly and all partners agree on that intention a joint intention emerges. Based on this joint intention joint actions can be planned and executed to reach a common goal.

2.2. *Action planning*

McDermott[43] defines planning as the part of the robot's program whose future execution the robot reasons about explicitly.

In planning it is typical to define a problem as a set of states with a set of possible associated actions. These actions may cause a transition from one state to another. Ideally a transition is caused from the current state to a desired state which is derived from joint intention. The general problem of planning is deciding upon a series of actions that will lead to reaching a goal, given the current state. Hence the peers need to know the state of the environment, of themselves and of the partners, as well as the abilities of all partners.

In real world problems the environment is not static and the outcome of actions and the transitions from one state to another may be uncertain. In addition, the state of the environment may not be known to a robot due to imperfect or insufficient sensors or partial observability of the environment. These complications have to be taken into account in planning for HRC. Some popular planning methods for real world scenarios are presented.

Decision theoretic planning[44] defines goal achievement through maximizing reward. Decision theoretic planning uses sequential decision models to decide on a sequence of actions to take in a series of states, rather than a one-time decision. There are sequential decision models that represent uncertainty, time-dependence, history dependence, and partial observability. However, decision theoretic planning is not suitable for reasoning about multiple partners with different goals. Multiple partners are combined into a model that presents the entire group as one single entity. In order to find a policy for only one partner acting in the presence of other partners, the other partners and their actions must be modelled as part of the state of the environment.

In **multi-body planning**[45] the partners of a team can either be modelled as a single body taking joint actions or as a game where all the players have common payoffs. Complications may arise when not all partners have the same knowledge about the state of the environment or each others' abilities and it is impossible to build an accurate centralized model. In human-robot interaction, only the robotic partners can be controlled, plans and actions of the human peers are not guaranteed. Still studies show that heterogeneous partners can assist each other in accomplishing their goals with only limited shared knowledge[46].

Other planning methods for multi-body planning include total-order planning[47] and partial-order planning[47,48]. In **total-order planning** a plan is produced from sub plans chronologically from a starting state or to an end state. It is defined as a search in the action space, as suitable actions are found chronologically for the sub problems. The linearization of the sub plans happens during the whole planning process. In **partial-order planning** sub problems are solved independently with sub plans, that are finally combined. Partial-order planning is defined as a search in the plan space, as first suitable plans are found for single sub problems. Linearization of actions happens in the end allowing for a higher flexibility. The advantage is flexibility in the order in which an action plan is constructed. Obvious and important sub goals can be planned first, rather than working in chronological order.

Independent of the planning strategy, plan libraries need to be created[49] first to have a suitable and sufficient set of actions available for given tasks.

Applications for planning in HRC are motion planning[50] and high level planning for complex cooperation[51]. Also safety aspects are very important in HRC and are being comprised in action planning[52]. Action planning needs still more resarch and combination with machine learning, to run more robustly and be able to react to new and unforeseen events.

Actions may have to be re-planned based on observations of the environment and intention of the peers. After planning actions that are supposed to lead to the desired goal, these actions are executed either by single partners or cooperatively.

2.3. *Joint action*

HRC is adopted in various fields and environments, thus actions and joint actions are manifold. Robots can move to target positions, fetch, hold, and support objects, carry them cooperatively with humans, speak or communicate in other ways, or search for something or someone. Some examples for robots, able to collaborate with humans are given below.

Leonardo[53] is a humanoid robot that is able to cooperatively solve a button pressing task, whilst communicating through gestures and understanding human speech and gestures. The mobile robot helper[30] is able to share a load and handle an object in collaboration with a human. Both Leonardo and Mobile Robot Helper are shown in Fig. 3 in collaboration with humans. RHINO[54] and its successor MINERVA[55] are museum tour guide robots communicating with and guiding museum visitors. The office robot Jijo-2[56] learns the locations of offices and persons by moving around and asking humans it encounters for information and directions. A robot developed by Kulyukin[57] is able to lead visually impaired people through structured indoor environments. Also cooperative sweeping[58] in a miniature setting has been researched. An overview of robots cooperating with other robots and humans is given by Fong[5].

Fig. 3. The robots Leonardo[53] and Mobile Robot Helper[30], collaborating with humans.

When robots are acting jointly with or close to humans safety is a crucial aspect. Dangers in actions might arise from robots or their arms when they have no or insufficient collision detection sensors. A robot that tips over might fall onto a human peer and injure him or her. Safety in HRC was a neglected research field for a long time but currently it is being contemplated[52,59].

2.4. *Machine learning*

Machine learning is a very large and complex research field. The methods of machine learning are complex and still developing. Due to the limited space, only a brief survey of machine learning is given here.

Often new and unforeseen situations arise in HRC, that can not all be programmed into a robot. Robots need the ability to adapt to unforeseen events, extend their knowledge and abilities, and learn new behaviours, actions and ways of estimating intent.

Machine learning algorithms are classified into supervised learning, unsupervised learning, reinforcement learning, learning to learn, and combinations of those. Supervised learning creates functions from training data, consisting of pairs of input objects and desired outputs. The task of the supervised learner is to predict values of functions for any valid input object after having been trained by a number of training examples. Unsupervised learning is distinguished from supervised learning by the fact that it models a set of input data and there is no a priori output. Reinforcement learning[60] is the closest to the human technique of learning. It is based on the principle of reward. Actions are taken so as to maximize some long-term reward. Reinforcement learning algorithms map states of the environment to actions the partners ought to take in those states. In learning to learn the algorithm learns its own inductive bias based on previous experience.

Machine learning has been widely investigated in recent works on cognitive robotics and intelligent agents. For gesture and speech recognition mostly pattern recognition techniques such as hidden Markov models[61] or artificial neural networks[20] are used. Pattern recognition systems are based on supervised learning and require a huge amount of training data. The approach of reinforcement learning in HRC[62] is to probabilistically explore states, actions, and outcomes, requiring a large amount of examples as well. Thus these techniques are often not appropriate in real collaborative environments where the number of examples may be small or may change quickly.

Recently classical approaches have been combined with psychological mechanisms like imitation, curiosity, selective attention, and memory. In new methods derived from nature, learning and teaching form a coupled system in which the learner and the teacher work together. Ideally, the teaching and learning process are closely coupled, well tuned, and transparent. Much attention has lately been given to the methods of robot imitation[63] and rhythm entrainment[64]. In robot imitation a robot learns motions and actions by imitating a human partner. In rhythm entrainment the imitation is taken one step further to perform a joint action synchronously with a human partner through evaluating frequency and phase shift. Other methods in HRC are imitative learning[65] and interactive teaching[66].

These machine learning techniques have applications in intention estimation, human-robot communication, action planning, interaction, and all other aspects of HRC.

3. Design of Collaborative Robots

Robots that are meant to collaborate with humans certainly require special abilities and design features, even more so when collaborating with fragile people such as children, elderly or disabled people.

The first thought while designing a robot for HRC must be on **safety**. A robot must not endanger a human under any circumstances. Since the very beginning of robotics this has been an evident fact and has been formulated by Isaac Asimov[67] in his famous Three Laws of Robotics. Safety issues have already been discussed in Section 2.3 for collaborative actions. Furthermore the hardware of the robot must not be dangerous for humans, banning sharp edges, points and accessible electric current. Danger might arise when a robot loses balance and falls, this has to be avoided through control and suitable hardware design.

Depending on the desired abilities and the purpose of the robot, software and hardware have to be carefully chosen. Possible software skills have been presented in Chapter 2. A humanoid hardware design is sensible when robots are supposed to collaborate naturally with humans. It could also have some features humans lack of, to supplement them. Depending on the robots task, sensors and actors may vary widely and have to be customized according to the area of application.

Thought has to be given also to the **appearance of robots** for HRC, as humans might allocate emotions to it and these emotions might effect the human performance. A robot supposed to collaborate with humans should not look threatening but friendly. Artificial skin of some kind helps to hide the metal skeleton, the actuators and wires which might look distressing to some humans. Actuators should not look dangerous, for example anthropomorphic hands will probably induce more of a feeling of safety than powerful looking gripper. A lot of thought is given to designing robot heads[68]. Depending on the purpose and character of the collaborations facial features can be included and the robot can look more or less similar to a human.

Mori[69] found a function that maps similarity to acceptance. He states that acceptance increases the more similar a robot becomes to a human, until a point is reached where the similarity is not quite 100% and subtle deviations from human appearance and behaviour create an unnerving effect. The ratio between humanness and machine-likeness is uncomfortable to humans at that point. The effect is called the **uncanny valley**, see Fig. 4.

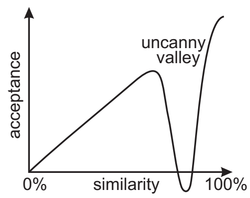

Fig. 4. Mori's function mapping acceptance to familiarity, with the uncanny valley[69].

This uncanny phenomenon is symptomatic of entities that elicit our model of human other but do not measure up to it. The human can obviously see that the peer is not human, but he or she doesn't know exactly what it is and what its abilities or possible dangers are. The class of very humanlike robots is often referred to as androids. An example for an android, is the ReplieeQ2[70], developed at Osaka University. The face of the ReplieeQ2, is an average composite face of several young Japanese women. The android has 42 air actuators for the upper torso, including 13 to realize facial expressions. The actuators enable unconscious movements, such as breathing and facial expressions, as well as conscious large movements of the head and arms.

Contrary to design policies that attempt to make robots more similar to humans a study by Yamada et al.[71] suggest that robots with typical robot appearence and a primitive expression of emotions are more easily understood by humans than more life-like appearing robots with complex ways of expressing emotions. This might be due to the present primitiveness of the cognition of robots. While a human like appearance, speech, and complex facial expressions give the impression of highly developed intelligence, a human is disappointed when finding that the robot is not as clever as it looks. Thus the robot's appearance and expressions should resemble its abilities.

Finally, another aspect that has to be taken into account, is the cultural environment a robot is supposed to work in. There are differences in the acceptance of robots by people with different cultural background[72]. Especially in Japan robots are accepted more positively than in European and American countries. Much of the research in robotics is done in Japan, and also the market for robots is already great there. Probably the acceptance of robots in the rest of the world will grow bringing along a high economic impact.

4. Applications for Human-Robot Collaboration

Applications for human-robot collaboration are manifold, the main application areas being healthcare, construction, space applications, search and rescue, tour guiding, and home service as well as entertainment. All of these areas have different requirements and the degree to which they are integrated varies.

Diverse **healthcare robots** are being developed, such as robots that guide the blind[57,73,74], robotic walkers[75] and wheelchairs[76], elderly care robots[77,78], and robots for the therapy of autistic children[79]. These robots are all designed to aid diseased or fragile humans and must be especially well designed and scrutinized. Ethical issues have to be considered particularly in the field of healthcare and medicine. Therefore in these areas few collaborative-robotic products are commercially available jet, though they are a major field of research.

A seminal field for HRC is **construction**, where robots can relieve human workmen of carrying heavy loads and ease repeating construction tasks. First steps toward joint construction have been taken by the mobile robot helper[30] that can share loads and handle objects in collaboration with a human. The JAST human-robot dialogue system[80] is capable of solving a construction task collaboratively with a human. Recently NASA has developed the robonaut[22] for **space construction applications**. Robonaut is currently teleoperated but may gain more autonomous abilities in the future. The advantage of having robots support humans in outer space is, that they are impassible towards the low temperatures and thus do not need long prearrangements before going into outer space. Also they can stay at the site, i.e. the space station, while human astronauts need to return to the earth and be replaced by others, which is very costly.

Another field, where robots support humans and take on dangerous tasks is **urban search and rescue**[81]. Robots are designed to move into collapsed buildings, collect data, and try and find human victims, who can then be rescued by human staff members. Robots for urban search and rescue need to be small, to get to inaccessible places. Also crawler tracks[82] or tires with a distinct tread pattern are major features as these robots have to move on uneven grounds. Favorable are shape-shifting robots[83], as they can go on even when they have tipped.

Some robots are already used as **tour guides** in Museums[55] and office environments[56,84]. They can guide humans to certain rooms or exhibitions and give information. Robotic guides may also in the future be deployed in shopping malls to show human visitors where to find what they are looking for, or as tour guides for tourists in city centers. Tour guide robots need the abilities to navigate safely through their specified human populated environment, and to communicate with humans explicitly.

Home service will be a major application area for HRC in the future. First robots have been developed that vacuum-clean and act as home security[85], registering intruders and capturing pictures of them. The household robot Wakamaru[86] developed by Mitsubishi is commercially available. It can connect itself to the network to provide necessary information for daily life, look after the house while the inhabitants are absent, and communicate. Truly collaborative home service robots are still a matter of research, such as a robot in an assistive kitchen[87], that can get plates and glasses from a cupboard and set the table.

Various **entertainment** robots capable of interaction are already commercially available. There are robotic pets[88] and humanoid robots[89,90] that communicate with humans and each other.

Though most of these systems are still a matter of research, it is only a matter of time until truly collaborative robots will be commercially available and part of our everyday lives.

5. Synopsis

HRC requires a common plan for all partners to fulfil a joint intention. The intentions of the other team members and what they are doing have to be known to gain a joint intention. Based on that knowledge robots can plan their own actions that will lead to reaching a common goal. Therefore, they need to be able to perceive and comprehend their environment, make decisions, plan their actions, learn, and reflect upon themselves and their environment. An architecture of the processes necessary for HRC has been presented.

A robot perceives the environment and the partners. Upon this information the robot can gain an understanding of the environment, estimate the intention of the partners, learn, and expand its knowledge. To collaborate the partners agree on a joint intention, derived from the single estimated intentions. The joint intention provides a common goal. Action planning is used to find a set of actions that will lead to that common goal. Finally joint actions are taken. The processes of intention estimation, action planning, joint action and machine learning have been reviewed. Additionally thoughts on robot hardware design have been given. Robots for HRC need to be safe and may under no circumstances endanger humans. Furthermore the robot's appearance can affect human emotions and needs to be carefully designed in order to gain acceptance.

The methods that make up HRC, come from the different research fields of cognitive sciences, artificial intelligence, mechanics, and psychology. While many researchers consider only a single aspect of HRC and research it thoroughly, they merely add other aspects if necessary. An all-embracing consideration of HRC and comprised methods is necessary to create really collaborative systems.

Acknowledgments

This work is supported in part within the DFG excellence initiative research cluster *Cognition for Technical Systems – CoTeSys*, see also `www.cotesys.org`.

References

1. J. Pineau, M. Montemerlo, M. Pollack, N. Roy, and S. Thrun, Towards robotic assistants in nursing homes: Challenges and results, *Robotics and Autonomous Systems*, **42**(3-4) (2003) pp. 271-281.
2. W. Bluethmann, R. O. Ambrose, M. A. Diftler, R. S. Askew, E. Huber, M. Goza, F. Rehnmark, C. Lovchik, and D. Magruder, Robonaut: A robot designed to work with humans in space, *Autonomous Robots*, **14**(2-3) (2003) pp. 179-197.
3. R. R. Murphy, D. Riddle, and E. Rasmussen, Robot-assisted medical reachback: A survey of how medical personnel expect to interact with rescue robots, in *IEEE International Workshop on Robot and Human Interactive Communication* (2004) pp. 301-306.
4. B. J. Grosz, Collaborative systems, in *AAAI Presidential Address* **2**(1996) pp. 67-85.
5. T. W. Fong, I. Nourbakhsh, and K. Dautenhahn, A survey of socially interactive robots: Concepts, design, and applications, *Tech. Rep. CMU-RI-TR-02-29* (Robotics Institute, Carnegie Mellon University, 2002).
6. S. Kiesler and P. Hinds, Introduction to this special issue on human-robot interaction, in *Human-Computer Interaction* **19** (2004) pp. 1-8.
7. H. A. Yanco and J. L. Drury, Classifying human-robot interaction: an updated taxonomy, in *IEEE International Conference on Systems, Man & Cybernetics* (2004) pp. 2841-2846.
8. J. R. Katzenbach and D. K. Smith, *The Wisdom of Teams: Creating the High-Performance Organization*, (New York: HarperBusiness, 1994).
9. P. R. Cohen and H. J. Levesque, Teamwork, *Tech. Rep. 504, Stanford University* (1991).
10. K. Davis, R. Biddulph, and S. Balashek, Automatic recognition of spoken digits, *Journal of the Acoustical Society of America* **24**(6) (1952) pp. 637-642.
11. D. B. Fry, Theoretical aspects of mechanical speech recognition, *Journal of the British Institution of Radio Engineers* **19**(4) (1959) pp. 211-218.
12. W. A. Woods, Understanding subsumption and taxonomy: A framework for progress, in *Principles of Semantic Networks: Explorations in the Representation of Knowledge* (1991) pp. 45-94.
13. R. Schank, Conceptual dependency: A theory of natural language understanding, *Cognitive Psychology***3** (1972) pp. 552-631.
14. R. Pieraccini and D. Lubensky, Spoken language communication with machines: the long and winding road from research to business, in *International conference on Innovations in Applied Artificial Intelligence* (Springer-Verlag, 2005) pp. 6-15.
15. M. You, C. Chen, J. Bu, J. Liu, and J. Tao, A hierarchical framework for speech emotion recognition, in *IEEE International Symposium on Industrial Electronics* **1** (2006) pp. 515-519.
16. P. Nugues, An overview of speech synthesis and recognition, in *An Introduction to Language Processing with Perl and Prolog* (Springer-Verlag, 2006) pp. 1-22.
17. M. Schröder, Emotional speech synthesis: A review, in *Eurospeech* **1** (2001) pp. 561-564.

18. Z. Li, N. Hofemann, J. Fritsch, and G. Sagerer, Hierarchical modeling and recognition of manipulative gesture, in *Workshop on Modeling People and Human Interaction* (2005).
19. V. Pavlović, R. Sharma, and T. S. Huang, Visual interpretation of hand gestures for human-computer interaction: A review, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(7) (1997) pp. 677-695.
20. K. Murakami and H. Taguchi, Gesture recognition using recurrent neural networks, in *SIGCHI conference on Human factors in computing systems* (ACM Press, 1991) pp. 237-242.
21. P. Maes, T. Darrell, B. Blumberg, and A. Pentland, The alive system: full-body interaction with autonomous agents, *Computer Animation*, (1995) pp. 11-18.
22. D. Kortenkamp, E. Huber, and R. P. Bonasso, Recognizing and interpreting gestures on a mobile robot, in *National Conference on Artificial Intelligence* **2** (1996) pp. 915-921.
23. T. Starner and A. Pentland, Visual recognition of american sigh language using hidden markov models, in *IEEE International Symposium on Computer Vision* (IEEE Press, 1995).
24. K. Sakita, K. Ogawara, and S. Murakami, Flexible cooperation between human and robot by interpreting human intention from gaze information, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2004) pp. 846-851.
25. T. Kanade, J. Cohn, and Y.-L. Tian, Comprehensive database for facial expression analysis, in *IEEE International Conference on Automatic Face and Gesture Recognition* (2000) pp. 46-53.
26. C. Breazeal and B. Scassellati, How to build robots that make friends and influence people, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (1999) pp. 858-863.
27. S. Sosnowski, A. Bittermann, K. Kühnlenz, and M. Buss, Design and evaluation of emotion-display eddie, in *IEEE International Conference on Robotics and Intelligent Systems* (2006) pp. 3113-3118.
28. Y. Hattori, H. Kozima, K. Komatani, T. Ogata, and H. G.Okuno, Robot gesture generation from environmental sounds using inter-modality mapping, in *International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems* 123, (2006), pp. 139-140.
29. K. Kosuge and N. Kazamura, Control of a robot handling an object in cooperation with a human, in *IEEE International Workshop on Robot and Human Communication* (1997) pp. 142-147.
30. K. Kosuge, M. Sato, and N. Kazamura, Mobile robot helper, in *IEEE International Conference of Robotics and Automation* (2000) pp. 583-588.
31. S. Gentry and R. Murray-Smith, Haptic dancing: human performance at haptic decoding with a vocabulary, in *IEEE International conference on Systems Man and Cybernetics* **4** (2003), pp. 3432-3437.
32. K. Kosuge, T. Hayashi, Y. Hirata, and R. Tobiyama, Dance partner robot – MS DanceR, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* **4** (2003) pp. 3459-3464.
33. Y. Kunii and H. Hashimoto, Tele-handshake using handshake device, in *IEEE/IECON 17 International Conference on Industrial Electronics*, Control, and Instrumentation, (1995) pp. 179-182.
34. D. Minnen, I. Essa, and T. Starner, Expectation grammars: leveraging high-level expectations for activity recognition, in *Computer Vision and Pattern Recognition* **2** (2003) pp. 626-632.

35. D. J. Moore, I. A. Essa, and M. H. Hayes, Exploiting human actions and object context for recognition tasks, in *IEEE International Conference on Computer Vision* **1** (1999) pp. 80-86.
36. B. Möller and S. Posch, Analysis of object interactions in dynamic scenes, in *DAGM Symposium on Pattern Recognition* (Springer-Verlag, 2002) pp. 361-369.
37. D. A. Baldwin and J. A. Baird, Discerning intentions in dynamic human action, *Trends in Cognitive Sciences* **5**(4) (2001) pp. 171-178.
38. A. Bobick, Movement, Activity, and Action: The Role of Knowledge in the Perception of Motion, *Royal Society Workshop on Knowledge-based Vision in Man and Machine* (1997) pp. 1257-1265.
39. M. Kranz, R. B. Rusu, A. Maldonado, M. Beetz, and A. Schmidt, A player/stage system for context-aware intelligent environments, in *System Support for Ubiquitous Computing Workshop, at the 8th Annual Conference on Ubiquitous Computing* (2006).
40. O. C. Schrempf, A. J. Schmid, U. D. Hanebeck, and H. Wrn, A novel approach to proactive human-robot cooperation, in *IEEE International Workshop on Robot and Human Interactive Communication* (2005) pp. 555-560.
41. D. Kulić and E. Croft, Estimating intent for human-robot interaction, in *International Conference on Advanced Robotics* (2003) pp. 810-815.
42. P. Rani, N. Sarkar, C. A. Smith, and L. D. Kirby, Anxiety detecting robotic system towards implicit human-robot collaboration, *Robotica* pp. **22**(1) (2004) pp. 85-95.
43. D. McDermott, Robot planning, *AI Magazine***13**(2) (1992) pp. 55-79.
44. J. Feldman and Y. Yakimovsky, Decision theory and artificial intelligence: I. a semantics-based region analyzer, *Journal of Artificial Intelligence***5**(4) (1974) pp. 349-371.
45. M. de Weerdt, A. ter Mors, and C. Witteveen, Multi-agent planning: An introduction to planning and coordination, in *European Agent Summer School* (2005) pp. 1-32.
46. J. H. Lawton and C. Domshlak, Multi-agent opportunistic planning and plan execution, in *IEEE International Conference on Tools with Artificial Intelligence* (IEEE Computer Society, 2004) pp. 408-415.
47. S. Minton, J. L. Bresina, and M. Drummond, Total-order and partial-order planning: A comparative analysis, *Journal of Artificial Intelligence Research***2** (1994) pp. 227-262.
48. X. Nguyen and S. Kambhampati, Reviving partial order planning, in *International Joint Conference on Artificial Intelligence* (2001) pp. 459-466.
49. A. Müller and M. Beetz, Designing and implementing a plan library for a simulated household robot, in *Workshop on Cognitive Robotics*(2006), pp. 119-128.
50. C. Esteves, G. Arechavaleta, and J. Laumond, Motion planning for human-robot interaction in manipulation tasks, in *IEEE International Conference on Mechatronics & Automation* (2005) pp. 1766-1771.
51. R. Alami, A. Clodic, V. Montreuil, E. A. Sisbot, and R. Chatila, Task planning for human-robot interaction, in *Joint conference on Smart objects and ambient intelligence* (ACM Press, 2005) pp. 81-85.
52. D. Kulić and E. A. Croft, Safe planning for human-robot interaction, *Journal of Robotic Systems* **22**(7) (2005) pp. 383-396.
53. C. Breazeal, A. Brooks, J. Gray, G. Hoffman, C. Kidd, H. Lee, J. Lieberman, A. Lockerd, and D. Mulanda, Tutelage and collaboration for humanoid robots, *International Journal of Humanoid Robots* **1**(2) (2004) pp. 315-348.
54. W. Burgard, A. B. Cremers, D. Fox, D. Hhnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, Experiences with an interactive museum tour-guide robot, *Artificial Intelligence* **114**(1-2) (1999) pp. 3-55.
55. S. Thrun, M. Bennewitz, W. Burgard, A. B. Cremers, F. Dellaert, D. Fox, D. Hahnel,

C. R. Rosenberg, N. Roy, J. Schulte, and D. Schulz, MINERVA: A tour-guide robot that learns, in *KI – Kunstliche Intelligenz* (1999) pp. 14-26.
56. T. Matsui, H. Asoh, F. Asano, T. Kurita, I. Hara, Y. Motomura, K. Itou, and J. Fry, Integration of Real-world Interaction Functions on the Jijo-2 Office Robot, *Bulletin of the Electrotechnical Laboratory* **64**(4-5) (2001) pp. 51–60.
57. V. A. Kulyukin, C. Gharpure, J. Nicholson, and G. Osborne, Robot-assisted wayfinding for the visually impaired in structured indoor environments, *Autonomous Robots* **21**(1) (2006) pp. 29-41.
58. K. Kobayashi and S. Yamada, Human-robot cooperative sweeping by extending commands embedded in actions, in *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2005) pp. 2977-2982.
59. V. Traver, A. del Pobil, and M. Perez-Francisco, Making service robots human-safe, in *International Conference on Intelligent Robots and Systems* **1** (2000) pp. 696-701.
60. L. Kaelbling, M. Littman, and A. Moore, Reinforcement learning: A survey, *Journal of Artificial Intelligence Research* **4** (1996) pp. 237-285.
61. L. R. Rabiner, A tutorial on hidden markov models and selected applications in speech recognition, *Proceedings of the IEEE* **77**(2) (1989) pp. 257-286.
62. Y. Wang, M. Huber, V. Papudes, and D. Cook, User-guided reinforcement learning of robot assistive tasks for an intelligent environment, in *International Conference on Intelligent Robots and Systems* ***1*** *(2003) pp. 424-429.*
63. M. Ito and J. Tani, On-line imitative interaction with a humanoid robot using a dynamic neural network model of a mirror system, Adaptive Behavior ***12****(2) (2004) pp. 93-115.*
64. Y. Maeda, A. Takahashi, T. Hara, and T. Arai, Human-robot cooperation with mechanical interaction based on rhythm entrainment – realization of cooperative rope turning, in IEEE International Conference on Robotics and Automation (2001) pp. 3477- 3482.
65. J. Demiris and G. Hayes, Imitative learning mechanisms in robots and humans, in 5th European Workshop on Learning Robots (1996) pp. 9-16.
66. J. Miura, K. Iwase, and Y. Shirai, Interactive teaching of a mobile robot, in IEEE International Conference on Robotics and Automation (2005) pp. 3378-3383.
67. I. Asimov, I, Robot (Gnome Press, 1968).
68. C. F. DiSalvo, F. Gemperle, J. Forlizzi, and S. B. Kiesler, All robots are not created equal: the design and perception of humanoid robot heads, in Symposium on Designing Interactive Systems (2002) pp. 321-326.
69. M. Mori, The uncanny valley, Energy ***7****(4) (1970) pp. 33-35.*
70. H. Ishiguro, Android science: Toward a new cross-interdisciplinary framework, in The Twelveth International Symposium of Robotics Research (2005).
71. S. Yamada and T. Komatsu, Designing simple and effective expression of robots primitive minds to a human, in IEEE/RSJ International Conference on Intelligent Robots and Systems (2006) pp. 2614-2619.
72. F. Kaplan, Who is afraid of the humanoid? investigating cultural differences in the acceptance of robots, International Journal of Humanoid Robotics ***1****(3) (2004) pp. 465-480.*
73. I. Ulrich and J. Borenstein, The guidecane – applying mobile robot technologies to assist the visually impaired, IEEE Transactions on Systems, Man, and Cybernetics ***31****(2) (2001) pp. 131–136.*
74. K. Iwatsuka, K. Yamamoto, and K. Kato, Development of a guide dog system for the blind with character recognition ability, in Canadian Conference on Computer and Robot Vision (IEEE Computer Society, 2004) pp. 401-405.

75. *D. Rodriguez-Losada, F. Matia, A. Jimenez, R. Galan, and G. Lacey, Implementing map based navigation in guido, the robotic smartwalker, in International Conference on Robotics and Automation (2005) pp. 3390-3395.*
76. *P. Jia and H. Hu, Head gesture based control of an intelligent wheelchair, in Conference of the Chinese Automation and Computing Society in the UK (2005) pp. 85-90.*
77. *M.Montemerlo, J. Pineau, N. Roy, S. Thrun, and V. Verma, Experiences with a mobile robotic elderly guide for the elderly, in National Conference on Artificial Intelligence (2002) pp. 587-592.*
78. *K. Wada and T. Shibata, Robot therapy in a care house – its sociopsychological and physiological effects on the residents, in IEEE International Conference on Robotics and Automation (2006) pp. 3966-3971.*
79. *I.Werry, K. Dautenhahn, B. Ogden, andW. Harwin, Investigating a robot as a therapy partner for children with autism, in European Conference for the Advancement of Assistgive Technology(2001).*
80. *M. E. Foster, T. By, M. Rickert, and A. Knoll, Symmetrical joint action in humanrobot dialogue, in RSS 2006 Workshop on Intuitive Human-Robot Interaction for Getting the Job done (2006).*
81. *R. R. Murphy, Human-robot interaction in rescue robotics, in Systems, Man, and Cybernetics (2004) pp. 138-153.*
82. *M. W. Kadous, R. K.-M. Sheh, and C. Sammut, Caster: A robot for urban search and rescue, in Australasian Conference on Robotics and Automation (2005).*
83. *B. Li, S. Ma, J. Liu, and Y. Wang, Development of a shape shifting robot for search and rescue, in IEEE International Workshop on Safety, Security and Rescue Robotics (2005) pp. 31-35.*
84. *E. Pacchierotti, H. Christensen, and P. Jensfelt, Design of an office guide robot for social interaction studies, in IEEE/RSJ International Conference on Intelligent Robots and Systems (2006).*
85. *B.-J. You, M. Hwangbo, S.-O. Lee, S.-R. Oh, Y. D. Kwon, and S. Lim, Development of a home service robot 'ISSAC', in IEEE/RSJ International Conference on Intelligent Robots and Systems (2003) pp. 2630-2635.*
86. *Wakamaru, Mitsubishi Heavy Ind. Ltd., www.mhi.co.jp/kobe/wakamaru/english/.*
87. *M. Beetz, J. Bandouch, A. Kirsch, A. Maldonado, A. Mller, and R. B. Rusu, The assistive kitchen – a demonstration scenario for cognitive technical systems, in COE Workshop on Human Adaptive Mechatronics (2007).*
88. *R. Sparrow, The march of the robot dogs, Ethics and Information Technology 4(4) (2002) pp. 305-318.*
89. *Y. Kuroki, M. Fujita, T. Ishida, K. Nagasaka, and J. Yamaguchi, A small biped entertainment robot exploring attractive applications, in IEEE International Conference on Robotics and Automation (2003) pp. 471-476.*
90. *Wowwee – Robosapien V2, http://www.robosapienv2online.com/.*

Andrea Bauer (*1981) is currently a research assistant (since December 2006) at the Institute of Automatic Control Engineering, Faculty of Electrical Engineering and Information Technology, Technische Universität München, Munich, Germany. She received the diploma engineer degree in Electrical Engineering in 2006 from the Technische Universität München, Munich, Germany. Her research interests include human-robot collaboration, multi-modal human-system interfaces, intention estimation, and automatic control.

Dirk Wollherr (*1975) is currently a postdoctoral researcher (since July 2005) at the Institute of Automatic Control Engineering, Faculty of Electrical Engineering and Information Technology, Technische Universität München, Munich, Germany. He received the diploma engineer degree in Electrical Engineering in 2000 and the Doctor of Engineering degree in Electrical Engineering in 2005 from the Technische Universität München, Munich, Germany. From 2001-2004 he has been research assistant at the Control Systems Group, Technische Universität Berlin, Germany. In 2004 he has been granted a research fellowship from the Japanese Society for the Promotion of Science (JSPS) at the Yoshihiko-Nakamura-Lab, The University of Tokyo, Japan. His research interests include automatic control, robotics, humanoid walking, human-robot interaction, and nonlinear control.

Martin Buss (*1965) is currently full professor (since June 2003), chair for control systems, and director of the Institute of Automatic Control Engineering, Faculty of Electrical Engineering and Information Technology, Technische Universität München, Munich, Germany. He received the diploma engineer degree in Electrical Engineering in 1990 from the Technical University Darmstadt, Germany, and the Doctor of Engineering degree in Electrical Engineering from the University of Tokyo, Japan, in 1994. In 1988 he was a research student at the Science University of Tokyo, Japan, for one year. As a postdoctoral researcher he stayed with the Department of Systems Engineering, Australian National University, Canberra, Australia, in 1994/5. From 1995-2000 he has been senior research assistant and lecturer at the Institute of Automatic Control Engineering, Department of Electrical Engineering and Information Technology, Technical University Munich, Germany where he finished his habilitation dissertation in the year 2000. From 2000-2003 he was a full professor and head of the Control Systems Group, Faculty of Electrical Engineering and Computer Science, Technical University Berlin, Berlin, Germany. His research interests include automatic control, mechatronics, robotics, intelligent control, multi-modal human-system interfaces, haptic (kinaesthetic, tactile, temperature) systems, medical haptic VR applications, optimization, nonlinear and hybrid discrete-continuous systems.