Proceedings of the Symposium

New Frontiers in Human-Robot Interaction

A symposium at the AISB 2009 Convention (6-9 April 2009)
Heriot-Watt University, Edinburgh, Scotland

Published by SSAISB: The Society for the Study of Artificial Intelligence
and the Simulation of Behaviour
http://www.aisb.org.uk/

ISBN - 190295680X

Introduction *New Frontiers in Human-Robot Interaction* Symposium

A two-day symposium at AISB 2009 (6-9 April 2009), Heriot-Watt University in Edinburgh, Scotland
http://homepages.feis.herts.ac.uk/~comqkd/HRI-AISB2009-Symposium.html

Held during the Science Festival (6-18 April 2009): http://www.sciencefestival.co.uk/

Symposium chair: Kerstin Dautenhahn (University of Hertfordshire, UK)

Motivation:

Human-Robot Interaction (HRI) is a growing research field with many application areas that could have a big impact not only economically, but also on the way we live and the kind of relationships we may develop with machines. Due to its interdisciplinary nature different views and approaches towards HRI need to be nurtured. This symposium will provide a platform to discuss collaboratively recent findings and challenges in HRI. Different categories of submissions are encouraged that reflect the different types of research studies that are being carried out. The symposium will encourage a diversity of views on HRI and different approaches taken. In the highly interdisciplinary research field of HRI, a peaceful dialogue among such approaches is expected to contribute to the synthesis of a body of knowledge that may help HRI sustain its creative inertia that has drawn to HRI during the past 10 years many researchers from HCI, robotics, psychology, the social sciences, and other fields.

Topics of interest include but are not limited to:

- Developments towards robot companions
- User-centred robot design
- Robots in personal care and health care
- Robots in search and rescue
- Sensors and interfaces for HRI
- Human-aware robot perception
- Dialogue and multi-modal human-robot interaction
- Robot architectures for socially intelligent robots
- HRI field studies in naturalistic environments
- Robot assisted therapy
- Robots in HRI collaborative scenarios
- Robots in schools and in other educational environments
- Robots as personal assistants and trainers
- Robot and human personality
- New methods and methodologies to carry out and analyze human-robot interaction
- Robots as companions and helpers in the home
- Robots as assistive technology
- Long-term or repeated interaction with robots

- Creating relationships with robots
- Expressiveness in robots
- Sustaining the engagement of users
- Personalizing robots and HRI interfaces
- Human-robot teaching
- Robots that learn socially and adapt to people
- User experience in HRI
- User needs and requirements for HRI
- Robots as autonomous companions
- Robots as remote-controlled tools
- Embodied interfaces for smart homes
- Ethnography and field studies
- Cross-cultural studies

The symposium encouraged submissions in any of the following categories.

**N* Completed empirical studies reporting novel research findings*
In this category we encourage submissions where a substantial body of findings has been accumulated based on precise research questions or hypotheses. Such studies are expected to fit within a particular experimental framework (e.g. using qualitative or quantitative evaluation techniques) and the reviewing of such papers will apply relevant (statistical and other) criteria accordingly. Findings of such studies should provide novel insights into human-robot interaction studies.

**E* Exploratory studies*
Exploratory studies are often necessary to pilot and fine-tune the methodological approach, procedures and measures. In a young research field such as HRI with novel applications and various robotic platforms, exploratory studies are also often required to derive a set of concrete research questions or hypothesis, in particular concerning issues where there is little related theoretical and experimental work. Although care must be taken in the interpretation of findings from such studies, they may highlight issues of great interest and relevance to peers.

**S* Case studies*
Due to the nature of many HRI studies, a large-scale quantitative approach is often neither feasible nor desirable. However, case study evaluation can provide meaningful findings if presented appropriately. Thus, case studies with only one participant, or a small group of participants, are encouraged if they are carried out and analyzed in sufficient depth.

**P* Position papers*
While categories N, E and S require reporting on HRI studies or experiments, position papers can be conceptual or theoretical, providing new interpretations of known results. Also, in this category we consider papers that present new ideas without having a complete study to report on. Papers in this category will be judged on the soundness of

the argument presented, the significance of the ideas and the interest to the HRI community.

**R* Replication of HRI studies*
To develop as a field, HRI findings obtained by one research group need to be replicated by other groups. Without any additional novel insights, such work is often not publishable. Within this category, authors will have the opportunity to report on studies that confirm or disconfirm findings from experiments that have already been reported in the literature. This category includes studies that report on negative findings.

**D* Live HRI Demonstrations*
Contributors may have an opportunity to provide live demonstrations (live or via Skype), pending the outcome of negotiations with the local organization team. The demo should highlight interesting features and insights into HRI. Purely entertaining demonstrations without significant research content are discouraged.

**Y* System Development*
Research in this category includes e.g. the design and development of new sensors, robot designs and algorithms for socially interactive robots. Extensive user studies are not necessarily required in this category.

Symposium contributions:

We invited unpublished, original work as extended abstracts (up to 3 pages) or full papers of up to 8 pages (double column). In category *D* we invited one page descriptions detailing the demo and its associated research questions.

Thirty-two submissions were received and 24 were accepted for presentation at the symposium. The symposium also includes one keynote invited talk as well as two invited talks. The symposium schedule emphasizes critical discussions of the presented research as well as wider issues that are important to HRI.

Acknowledgements:

The symposium organization would not have been possible with the dedicated supported of the international Programme Committee as well as the support of the AISB 2009 Convention Chair Dr. Nick Taylor and his Organization Committee.

Invited Talks

Invited Keynote Talk

Speaker: Holly Yanco (University of Massachusetts Lowell, USA)

Title: How to Partner with a Robot: The Design Space of Human-Robot Interaction

Abstract:
Assistive robots, such as robotic wheelchairs or wheelchair-mounted robotic arms operate in the same space as their user. Other systems, such as those designed for urban search and rescue, are located at a distance from their operator. Distance is just one of the many parameters that impact how we must design for effective human-robot interaction. For example, there can be a wide range of abilities in the target population for a robot system, particularly for assistive robots. This talk will present several dimensions that must be considered in the design of human-robot interaction and how to address them. Examples of robot systems will be drawn from many application domains.

Short invited Talk

Speaker: Takanori Shibata (National Institute of Advanced Industrial Science and Technology, Japan)

Title: Integration of Therapeutic Robot, Paro, into Elderly Care in Denmark

Abstract:
Since 1993, Paro, a baby seal robot, has been developed for two purposes: one is for as companion at home, and the other is for therapy at hospitals, elderly institutions, schools, and so on. In Denmark, Danish Technological Institute (DTI) have been distributing Paro to elderly institutions in Denmark since late 2008. So far, about 50 elderly institutions out of 1,200 in Denmark have been using Paro for elderly care, especially for caring elderly with dementia. DTI plan that they will introduce 1,000 Paros to elderly institutions in Denmark by 2011.

Short invited Talk

Speaker: Ruth Aylett (Heriot-Watt University, Scotland)

Title: Affect, empathy and graphical characters

Abstract:
Is it possible to produce interactive graphical characters about which the user really cares? How far are we from creating an 'affective loop' between user and character? Is this a requirement for companionship or not? The talk will briefly survey these questions and finally discuss how far work in graphical characters is applicable to robots.

New Frontiers in Human-Robot Interaction

A two-day Symposium as part of the AISB 2009 convention, 8-9 April 2009, Edinburgh, Scotland

Table of Contents:

Proceedings New Frontiers in Human-Robot Interaction, K. Dautenhahn (Ed.), symposium at the AISB09 convention, 8-9 April, Edinburgh, Scotland

An Architecture Supporting Proactive Robot Companion Behavior

Muhammad ALI and **Samir ALILI** and **Matthieu WARNIER,** and **Rachid ALAMI**[1 2]

Abstract. An essential aspect of human robot interaction is proactive robot behavior particularly in situations where the robot is able to determine by itself if, how and when it can intervene and help. This is certainly valuable since it permits the user to be freed from the burden of permanently monitoring the robot and choosing the command that should be issued to the robot. In this work we present an architecture for proactive robot behavior. Its main features involve the ability to select high level goals based on scenario recognition. The goals are then refined by a specific planner that is able to determine if the robot can contribute to the goal achievement and finally a human aware supervision system that allows the robot to share the human activity thanks to its ability to achieve task cooperatively. The paper describes the overall system and its implementation on a realistic testbed.

1 INTRODUCTION

When we see an elderly person moving a table by itself, we take initiative and take the other side of the table and help him move that table. Similarly robot companions in human environment will need to take initiative and help their human partner without requiring an explicit command. This would lead to a fluent human robot interaction and ultimately enable robot to create strong bonds with the human (one of the important challenge in human robot interaction [33]).

For robot to take initiative it should be able to answer following questions: What to observe? When to take initiative? How to insert itself to the on-going activity? For the first question, robot needs to have an recognition mechanism to detect scenarios (situations) calling for robot action, like: to help the elderly person moving a table. Second one will require a robot to take decision whether to be proactive or not: Does it have the capacity to hold the table? Finally a mechanism to hold current robot goals (stop cleaning the room and redo it later, for instance) and execute and supervise the new course of decided actions: Grab the table and also keep pace with human (ability to do joint task).

The work presented here consists of a system that permits proactive robot behavior by addressing a number of above mentioned issues. It focuses on (1) the use of chronicle recognition mechanism that can efficiently detect activities to which to robot may potentially participate, (2) a human aware planner that is able to determine if, how and when the robot can contribute and (3) a supervision system which is responsible for monitoring and executing robot tasks in close cooperation with human partners.

[1] CNRS ; LAAS ; 7 avenue du colonel Roche, F-31077 Toulouse, France, email: firstname.lastname@laas.fr

[2] Université de Toulouse ; UPS, INSA, INP, ISAE ; LAAS ; F-31077 Toulouse, France.

Section 2 discusses related work. Section 3 presents the system architecture and then, sections 3.1, 3.2, 3.3 and 3.4 give a short description of the components involved in the decisional layer. Section 4 deals with the implemented system and illustrates its use through several examples involving a mobile robot behaving proactively. Finally, section 5 concludes and discusses future work.

2 RELATED WORK

Work related to proactive robot behavior initially began with mixed initiative approaches. In mixed-initiative approach focus is on initiative shifts between human and robot, and is related to robot tasks. Mixed-initiative (also called facilitated initiative [20]) based on operator modalities [15][16] use a control architecture that allows robot to have different levels of autonomy. It can be in tele-operated, safe mode, shared control, collaborative task mode (CTM) and totally autonomous mode. Robot can take varying degree of "initiative" based on the mode chosen, the current context and even the difficulty of the task at hand. For example robot takes initiative and leads in navigation tasks in CTM mode.

A planner based mixed initiative approach is used in search and rescue scenario by[5]. Its architecture is based on model based execution monitoring (activities model defined) and a reactive planner monitors task execution using that model. If human operator changes execution order, planner responds by proposing a new execution order to him.

[3] uses an affect based mixed-initiative interaction approach using human robot interface. Robot responding to changes in human operators emotions (detecting drowsiness,inattentiveness etc) can take initiative from or may offer it back to human. Some approaches[2] also use emotion based planning for mixed initiative interaction.

Other methods like [29], use initiative for removing ambiguity in human intentions. Architecture consists of intention recognition using Dynamic Bayesian Networks and planner for task execution. Planner executes robot tasks for correctly inferred intentions and for ambiguous intentions planner selects an action from a table (defined by human) to induce human response and remove ambguity. In robot care [9][8], robots shows proactive behavior based on activity monitoring using activities defined as a schedule. And constraint violations of schedule trigger system initiative and perform some action in the form of a alarm or suggestions to the assisted person.

In our context, proactive behavior is not only based on constraint violations[9],or governed by operator modality[15][16] or planning based[5][29] but as our system aims at multi-layered proactive behavior it consists of a whole system from detection of robot goals (for itself and as well as for human goal), their management through task agenda, a planning mechanism that plans taking into account human and finally an execution and monitoring system takes into account human at every level of interaction. Affect based [3][2] approaches would be suited for initiative taking in close human robot contact situations.

3 SYSTEM ARCHITECTURE

We have devised a control architecture dedicated to robot decision and action in a human context (Figure 1). It enhances robot capacity to be proactive for robot related as well as human related tasks.It has been developed as an instance of the generic the LAAS Architecture for Autonomous Systems[4].

The decisional layer consists of four components:

- The **Task Agenda** A mechanism for robot higher level goal management.
- **CRS** An chronicle recognition system for modeling and recognizing scenarios.
- **HATP**: A task planning system that is able to synthesize socially acceptable robot plans that may involve human-robot collaborative action.
- **SHARY** which constitutes the decisional kernel. It is based on an incremental context-based task refinement in a human context.

We will now explain the different parts of this architecture.

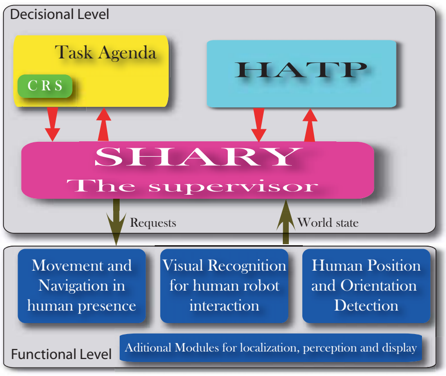


Figure 1. Architecture for proactive robots companions: The decisional layer consists of four components. SHARY, which is in charge of task supervision and execution. HATP the planner which provides other decisional abilities, and Task Agenda which manages high-level robot goals in a human and CRS is in charge of the interpretation activity of the persons in the robot vicinity.

3.1 Task Agenda

The role of the Task Agenda is to manage high-level robot goals and their associated tasks. It maintains an ordered list of high-level tasks and is embedded with basic mechanism to pre-empt and suspend current task if a higher priority task arrives and reschedule it when higher priority tasks end. New Agenda tasks are generated either by users requests (through multi-modal dialog) or on chronicles recognized by CRS. Currently task priorities are staticaly set according to human involvement in task initiation and execution. These priorities are not modified by plan modification and interaction history. Similarly task relevance is not tested when resuming execution. This should be improved soon.

3.2 CRS

A robot companion besides acting on human commands needs to actively monitor its environment and infer new goals for itself and human partners goals. It needs to incorporate a capacity to recognize an activity happening in its vicinity. Our approach is to model these events into temporally constrained networks called chronicles[23] (representing a situation or scenario)and using Chronicle Recognition System (CRS)[18] (developed by [17]) for monitoring these scenarios.

For example breakfast cooking scenario: First we can define events for this chronicle, human takes frying pan from kitchen-shelf, gets eggs from the fridge and turns on stove. Then we need to define temporal constraints on these events, these events can happen between 3 - 5 minutes and these events should occur in the morning. When recognized, robot can start setting the table for the breakfast while the human cooks.

We have defined many scenarios based on presence or absence of human activity around robot. These scenarios are so modeled to enable system to have different proactive behaviors. For example robot can look to find goals for itself (do housekeeping, be curious etc, like[30]) or find a human goal where it can help (like setting the breakfast table for human). When a chronicle model matches a defined scenario, CRS informs the supervisor. Supervisor acts according to behavior it represents, can add it as a new robot task in task agenda or if relevant to human goal, requests the planner for a plan.

The key aspects for CRS use are:

- It handles time explicitly.
- It can monitor modeled scenarios on the fly, by matching observations to model events and temporal constraints propagation.
- It can maintain several hypothesis (chronicles), that is, complete tree of instances of partial chronicles currently taking place.
- Can easily integrate several events in different scenarios and keep their window of relevance with respect to each scenario.

Human activity is essentially difficult to model, different humans will approach an activity differntly causing uncertainty around scenario recognition, for example, in a cooking scenario, there can be several different ways a human can start the activity, can take the eggs from fridge, can get frying pan from kitchen-shelf or can turn on stove first etc. One way to handle this uncertainty is by defining several chronicle for a same situation. Complexity does not increase significantly due to multiple chronicles as CRS is quite efficient in handling many chronicle insances[18].

3.3 Human Aware Task Planner - HATP

HATP is a planner designed for heterogeneous Agent interactions, in our case humans and robots. It is based on hierarchical task planning[24] and integration of behavior rules, which orient robot decision and produce social plans. HATP has also its own language[6], which allows us to model human preference, ability and capacity as we can see in the figure 2 we describe the fact that the human needs Glasses by a boolean attribute associate to the entity “Human” called "needGlasses" it has the true value if human needs the glasses and false otherwise. We complement the human model with the action description which takes into account the fact that the action is performed by the human or by the robot.

We take inspiration from human interaction to establish rules for a right social behaviors in human robot interaction. We define six types of rules[6]:

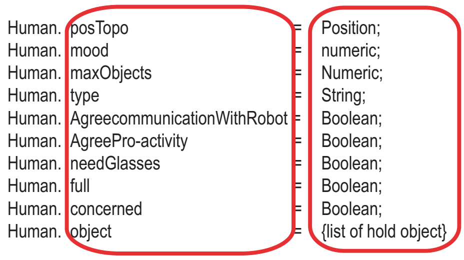


Figure 2. HATP Human model: In this example we can see HATP building the human model. The entity "HUMAN" is described by a set of attributes which represents respectively human topological position, degree of human desire to be involved in task, human ability, boolean attribute describing if human allow robot pro-active behaviour, boolean attribute describing if human needs glasses, boolean attribute describing if human is concerned about current task, list of objects that human takes.

- Undesirable states
- Undesirable sequences
- Bad decompositions
- Effort balancing
- Timeouts
- Crossed links

HATP planning process is composed of two threads. One thread is responsible for the plan refinement[28] and a second thread is responsible for plan evaluation. The second one is based on the Analytic Hierarchy Process (AHP)[21], it gives to the plan evaluation a total control on the plan quality because it combines the penalty added by the rules violations with the costs of actions. Both of them integrate human model. For example we can model the human desire to be involved in the task and the fact that he/she has physical handicap. In this situation HATP will produce plans involving the human in the task that respect his/her capability. Otherwise it produces plans with as least as possible human involvement.

In this paper, the main HATP performances is its ability to take into account human actions and to produce social plans, its capacity to handle contingencies, and also the possibility for HATP to start planning from a partial plans[6] (It gives the robot the possibility to analyse human plans and correct or complement them for proactive behaviour).

3.4 SHARY : The supervision and execution system

SHARY'S[11], [12]originality, as a supervision system, lies in its ability to take into account not only the task achievement but also communication and monitoring needed to support interactive task achievement in a flexible way. SHARY allows to define a task or a hierarchy of tasks linked to more or less elaborated "communication policies" that enable to execute tasks given the possibility to deal with contingencies that could occur during its execution (or even to stop the task in case of unexpected events).

A communication scheme, for a given joint task, represents all possible turn taking steps and synchronization between the robot and its human partner [13]. Each time a state is visited the corresponding task recipe or atomic task is launched.

From a practical point of view, a communication scheme is a finite state automaton. Its states are communication acts expressed by the robot through dialog or by an expressive motion. Its transitions are communication acts directly expressed by the human or inferred from her/his behavior by monitoring tasks.

We have some generic communication scheme with a defined set of communication acts that are mandatory in the framework of task achievement [1]. This set takes inspiration from Joint Intention Theory ([14]) that states that each partner should be informed of the beginning, realization and ending of a joint task.

While executing a specific task this generic communication acts will be instantiated as an act_X_task with a recipe, an execution state, etc. For example, when the robot is proposing to give the human an object, it is realizing the act_X_task defined by the `Give Object` task and the `ASK-TASK` act.

Task Recipe: Task recipes are methods that compute the partially ordered list of subtasks of an act_X_task. This sub-task tree contains both a set of tasks needed for the achievement of the act_X_task but also a list of tasks required for monitoring the execution. Recipes can be scripts, i.e. provided by the programmer, or can be synthesized by a planner such as HATP [28] presented previously.

Figure 3 describes SHARY execution at a given task level and exhibits the incremental context-based task refinement process which results in a dynamic hierarchical task tree.

Toward Proactivity: Using the system proposed so far we already managed to make the robot create proactively new robot only tasks (like cleaning a table) and also to ask human for help proactively if robot couldn't achieve one of its task alone (cannot pick up a bottle). We improve the system to make the robot act proactively for human goals recognized and instantiated by CRS.

If the new human goal recognized by CRS has a higher priority than ongoing task, the robot starts managing the human goal as if it has been directly asked by human until it receives the recipe from HATP. We then add a check step to make the robot monitor if the recipe given by HATP contains robot sub tasks and stop the task otherwise. Finally the robot asks human permission to realize the action described in the recipe and continues execution only if human agrees. The first step prevents the robot from bothering the human for nothing. The second steps ensure human agreement with robot initiative.

4 EXPERIMENTS

The system has been implemented and integrated on a fully equipped mobile manipulator called Jido and tested in the experimental environment, shown in figure4. It simulates a living-room with different objects of everyday human life (tables, chairs,. . .). The robot goal is to assists human in his daily life, for this the robot able to perform a number of tasks Serving, Cleaning and maintaining an updated knowledge of the state of the world (detecting and tracking persons in the room, detecting and recognizing new objects placed on tables by the persons. . . .).

Besides decisional capabilities that we have discussed, our robot has its functional level rich set of capabilities, as that involve motion[25], navigation[32], manipulation[31], perspective placement [27] in presence of humans as well as perception of human activities[7] and face tracking[22]. The challenge for the system is to perform these tasks showing proactive robot behavior and also interleaving it with social behavior when interacting with the human. We present, in the sequel, illustrative examples of Jido capabilities. A human living room furnished environment (Coffee Table,Cupboard Table chairs, Cupboard, Chairs) and objects (Glasses for reading, Books, Bottles, 2 Glasses for drink) is the setting for the examples.

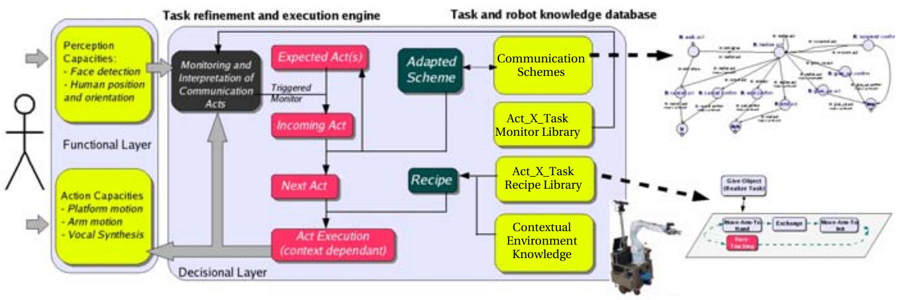
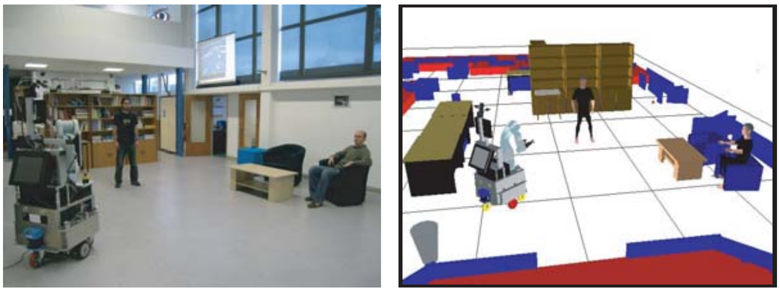


Figure 3. General Description of Shary (at a given task level inside a hierarchy of tasks): when the task is created, a communication scheme associated to the task is instantiated according to the task, the context and the concerned agent = *Adapted Scheme*. This scheme gives the first act to execute. The recipe corresponding to that act (precisely to this act_X_task) is instantiated by the help of a recipes library: *Recipe*. During *Act Execution*, communication and execution monitoring is done through wait on *Expected Acts*. When a monitor is triggered *Incoming Act*, i.e. when an expected act happens, the current act is stopped and the answer is instantiated given the communication scheme *Next Act*. And so on...



(a) A "living-room" (b) 3D model used by the robot

Figure 4. Robot working environment.

The examples will focus on three different kind of proactive robot behavior:

Example 1: proactively generate new tasks involving robot

As mentioned in section 3.1 the Task Agenda implements the ability for the robot to manage its high-level goals. New tasks can be obtained via dialog or can be generated via CRS based on the recognition of chronicles representing human activity[18]. In the task Agenda there are some modes dedicated to the achievement of a family of tasks. These modes are defined such as "clean", "serve" and also "Curiosity". While in this mode the robot will try to achieve tasks to keep an updated world model of objects on the table if the robot had inferred possible changes through CRS.

In this example human comes near the table and stays there for some time and leaves. When this chronicle is recognized, CRS sends an Update Knowledge request to the supervisor indicating uncertainty of the table state. And when robot is not doing a more prioritized task it goes and looks at the table to find current table state. This helps robot keep an up-to-date world model of objects on the table. Figure 8 illustrates the example described.

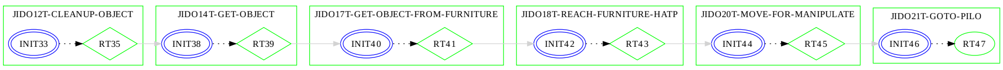
Example 2: proactively ask human help to achieve a robot task

Through this scenario, we would like to show the ability of the system to take initiative and ask human help to remove an ambiguity or to escape from a blocked situation. The task consists in Jido cleaning the living-room table by picking up the bottle and throwing it into the trash. Figure 6 illustrates the plan produced by HATP for normal execution as well as a snapshot of a current task refinement decomposition performed by SHARY. Initially HATP produces a plan where Jido can do the task itself. SHARY executes the plan. During the execution of the (move to table) task our human aware motion planner (MHP[31]) places our robot in front of the table for safe manipulation in human presence. Sometimes Jido cannot reach the bottle. SHARY would then ask HATP to replan. If there is a person present in the environment that can reach the bottle, HATP will provide a new plan where human will reach the bottle and give it to Jido. Executing this plan will consist in Jido asking the human to give it the bottle.

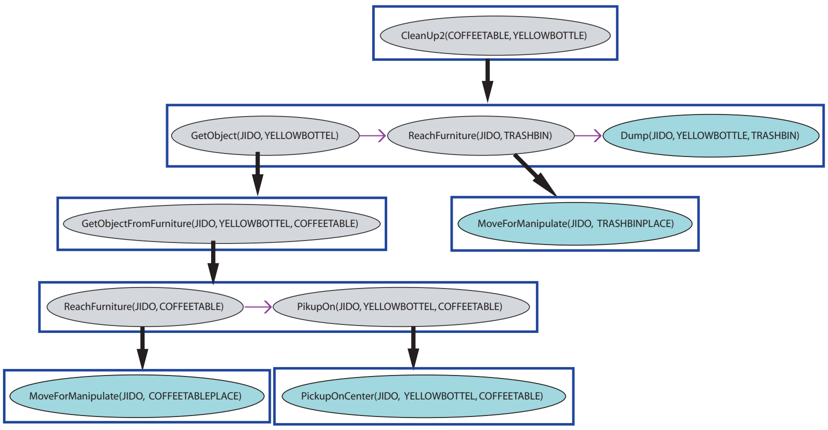
Figure 7 illustrates the different steps including request to the planner when robot can ask human help.

(a) (b) (c) (d) (e)

Figure 5. Example 3: proactively generate new tasks involving robot. a person approaches the table near the cupboard and stays still for a moment before leaving. This induces the fact that the person might have put or taken bottles. Jido takes the initiative to approach the table and to update its knowledge using its perception functions.



(a) Current execution task stack in SHARY: Boxes are tasks, circles and diamonds shapes are act_X_tasks (`RT` is the abbreviation of `REALIZE-TASK`), gray arrows represent decomposition links and dotted arrows are transitions between act_X_tasks inside a communication scheme. Blue color corresponds to *achieved* tasks and acts while green color means that they are being executed.



(b) Hierarchical plan from HATP for the `clean-up` task

(c) Snapshot from the experiment

Figure 6. Example 2: proactively ask human help to achieve a robot tasks Achieving `clear-up(yellow-bottle)` consists mainly in achieving its act_X_task `RT` or its plan HATP as well. The HATP plan stops at a given abstract level in task decomposition (6(b)). Consequently, SHARY needs to further refine these tasks corresponding to the leaves in the HATP plan tree. This is illustrated in 6(a) for `MoveForManipulate` task.

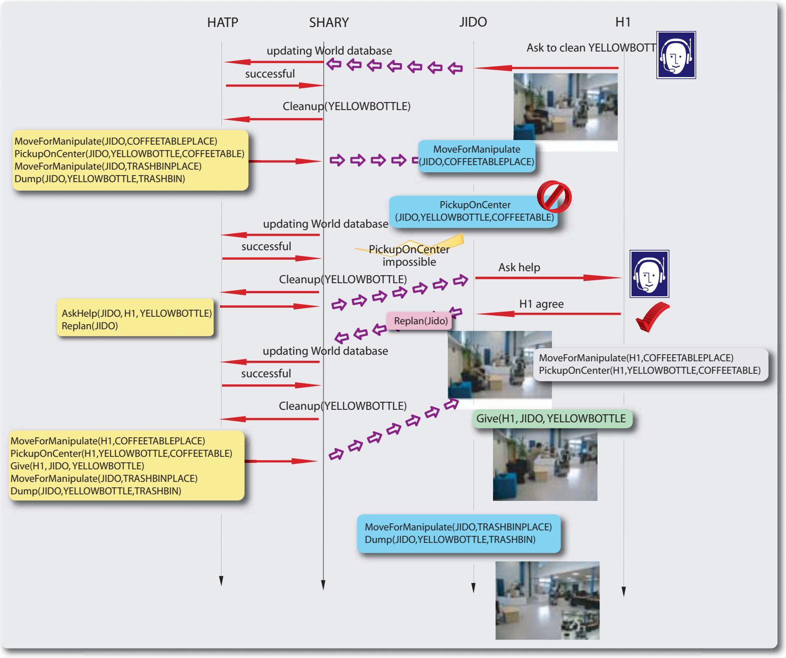


Figure 7. **Example 2(In difficulty ask Help):** `clean-up` task execution: At the top left of the figure, we see a simple version of the first HATP plan computed to achieve the `clean-up`. In the middle and at the right side of the figure, we see the execution stream corresponding to this plan execution. This first plan failed due to robot inability to take the bottle (even when it has perceived it). SHARY asks a new feasible plan. HATP finds a plan with a higher cost and two streams and where the person is requested to participate by giving the bottle to Jido. The robot can then proceed and move to throw the bottle in the trash bin.

Example 3 : act proactively to help human achieve his goal

In this scenario we will see the capacity of the system to generate a pro-active behavior to help human achieve his/her goal. Robot observes the environment via CRS and if it recognizes a scenario that corresponds to human goal, it transmits this human goal to SHARY. SHARY adds this probable new goal in the task Agenda which analyzes its priority in comparison to the other present in the TODO list. If the new goal has the priority SHARY requests HATP for a plan, HATP searches for plans that minimize human effort. If it exists at least one, it supplies the best one to the SHARY. If the plan contains some task for the robot, Shary starts the execution otherwise there is no thing to do for the robot and it carries on its current activity.

In this example, Jido observes human going near the bookshelf, taking the book and sitting on the sofa. CRS recognizes this scenario as human wants to read and informs SHARY. Which adds this task to the Agenda which puts it on the top of the list after analysis. Shary sends a planning request to HATP. HATP starts planning, taking into account human preference, ability and capacity (like John wears glasses where as Jack does not) finds a plan where human should have his reading glasses (Assuming an ambient camera system through which jido knows that human is not wearing glasses) because one of the precondition of the **to read** task is to have reading glasses. SHARY checks if there is some thing to do and executes HATP plan if there is. Figure 8(a) shows the system activities flow for this example. Figure 8(b) shows, the HATP plan with two streams, one for human, showing his course of action and other strean showing robot course of action to intervene and help human achieve his goal of reading by bringing the reading glasses for him.

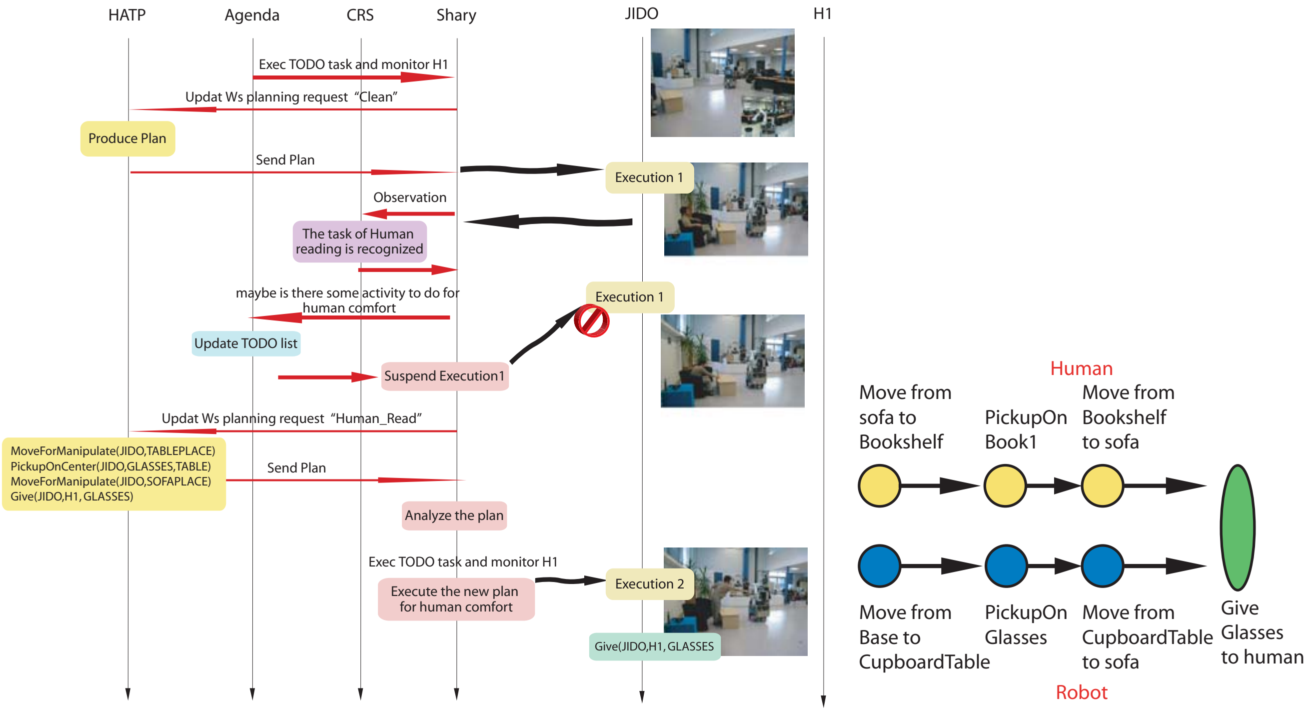
CONCLUSION AND FUTURE WORK

In this paper, we have presented a system architecture for proactive robot initiative taking, adding to its already rich set of human robot interaction (hri) capabilities. We have discussed main components involved in initiative taking and describe how they are working in our system. We have demonstrated through implemented examples various aspects of the system, scenario recogniton, planning based initiative taking and execution through hri dedicated supervisor.

Deciding on whether to take initiative and whether to ask permission or inform about initiative taking is not easy. These issues were simplified here. The robot always took initiative if there are some plan involving robot and always asked for permission before execution.

Human preferences regarding robot initiative taking were not taken into account to produce plan. Some people could be reluctant to initiative taking robots whereas some other would be very enthusiastic and possibly bothered that the robot asks permission each time it wants to take initiative. We could also imagine that the individual preference of a particular person evolves both in a very short term (Emergency situations, Human emotional state, recent dialogues and actions) and in a longer term (people gradually becoming more confident or conversely more suspicious or annoyed). HATP is able to model human preferences so we could very easily introduce parameters stating individual preferences toward robot productivity to adapt the resulting plans to each individuals. The new challenge would then be to make this preference evolves according to context and interaction memory.

Concerning scenario recognition, we can use chronicle learning[26] for obtaining new chronicles or focus on proba-



(a) While the robot is doing some task, CRS detects a human reading chronicle. It informs SHARY that create a task in the task Agenda. SHARY sends a planning request to HATP, HATP produces a plan. SHARY analyses the plan and if there is a robot task for helping human executes it.

(b) HATP produced plan, with two streams: one for human course of action, and another for robot actions where it can proactively intervene to help human

Figure 8. Example 3: act proactively to achieve human goal The general flow of activities in the system

bilistic approaches for better handling of uncertainty. Or can look to probabilistic approaches for interpreting human activity, like[10, 19].

Task agenda being based on fixed task priority prohibits natural and rational behavior. It needs to be more dynamic, should take into many factors, for instance, task progress. For that plan monitoring will be important, CRS can be useful if plans can be synthesized into chronicles.

Acknowledgments: The work described here has been partially conducted within the EU Integrated Project CHRIS (Cooperative Human Robot Interaction Systems) funded by the E.C. Division `FP7-IST` under Contract `215805`. We also aknowldege support of Agence Nationle de la Recherche ANR (AMORCES/PSIROB project). Finally we would also like to thank Christophe Dousson of Research and Development Department at France Telecom for providing CRS. 1st Author acknowledges the Higher Education Commission (HEC) of Pakistan and Government of France for their financial support during his studies.

REFERENCES

[1] A.Clodic, R.Alami, V.Montreuil, , and et al, 'A study of interaction between dialog and decision for human-robot collaborative task achievement', in *16th IEEE RO-MAN*, (2007).

[2] MariaJose Acosta, Dongyeop Kang, and Ho-Jin Choi, 'Robot with emotion for triggering mixed-initiative interaction planning', in *CIT-WORKSHOPS '08: Proceedings of the 2008 IEEE 8th International Conference on Computer and Information Technology Workshops*, pp. 98–103, Washington, DC, USA, (2008). IEEE Computer Society.

[3] J. Adams, P. Rani, and N. Sarkar, 'Mixed initiative interaction and robotic systems', in *Workshop on Supervisory Control of Learning and Adaptive Systems, Nineteenth National Conference on Artificial Intelligence (AAAI-04)*, San Jose, CA, USA, (2004).

[4] R. Alami, R. Chatila, S. Fleury, M. Ghallab, and F. Ingrand, 'An architecture for autonomy', *Internatonal Journal of robotics Research, Special Issue on Integrated Architectures for Robot Control and Programming*, **17**(4), (1998).

[5] Finzi Alberto and Orlandini Andrea, 'Human-robot interaction through mixed-initiative planning for rescue and search rovers', in *AIIA '05: Proceedings of the 9th Congress of the Italian Association for Artificial Intelligence on AIIA 2005*, pp. 483–494, Berlin, Heidelberg, (2005). Springer-Verlag.

[6] Samir Alili, Rachid Alami, and Vincent Montreuil, 'A task planner for an autonomous social robot', in *The 9th International Symposium on Distributed Autonomous Robotic Systems 2008 (DARS2008) in Tsukuba International Congress Center, Tsukuba, Ibaraki, Japan*, (November 17-19, 2008).

[7] B. Burger, I. Ferrane, and F. Lerasle, 'Multimodal interaction abilities for a robot companion', in *ICVS*, (2008).

[8] Amedeo Cesta, Gabriella Cortellessa, Federico Pecora, and Riccardo Rasconi, 'Supporting interaction in the robocare intelligent assistive environment', in *AAAI '07: In Proceedings of AAAI Spring Symposium on Interaction Challenges for Intelligent Assistants 2007*, pp. 18–25, (2007).

[9] Amedeo Cesta, Gabriella Cortellessa, Federico Pecora, and Riccardo Rasconi, 'Synthesizing proactive assistance with heterogeneous agents', in *AIIA '07: Proceedings of the 10th Congress of the Italian Association for Artificial Intelligence on AIIA 2007*, pp. 495–506, Berlin, Heidelberg, (2007). Springer-Verlag.

[10] Pau-Choo Chung and Chin-De Liu, 'A daily behavior enabled hidden markov model for human behavior understanding', *Pattern Recogn.*, **41**(5), 1589–1597, (2008).

[11] A. Clodic, *Supervision pour un robot interactif : Action et Interaction pour un robot autonome en environnement humain*, Ph.D. dissertation, University of Toulouse, 2007.

[12] A. Clodic, H. Cao, S. Alili, V. Montreuil, R. Alami, and R. Chatila, 'Shary: a supervision system adapted to human-robot interaction', *11th International Symposium on Experimental Robotics 2008, ISER 2008*, (2008).

[13] A. Clodic, M. Ransan, R. Alami, and V. Montreuil, 'A management of mutual belief for human-robot interaction', in *IEEE SMC*, (2007).

[14] P. R. Cohen and H. J. Levesque, 'Teamwork', *Nous*, **25(4)**, 487–512, (1991).

[15] Nielsen Curtis W., Bruemmer David J., Few Douglas A., and Walton Miles C., 'Mixed-initiative interactions for mobile robot search', in *Proceedings, The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference, July 16-20, 2006, Boston, Massachusetts, USA*. AAAI Press, (2006).

[16] Bruemmer D. J., Few D. A., Nielsen C. W., and Walton M. C., 'Mixed-initiative control for collaborative countermine operations', in *IEEE Transactions on Robotics*. IEEE Robotics and Automation Society, (2007).

[17] C. Dousson, 'Alarm driven supervision for telecommunication network : Ii- on-line chronicle recognition', *Annals of Telecommunication*, 501–508, (1996).

[18] Christophe Dousson and Pierre Le Maigat, 'Alarm driven supervision for telecommunication network : Ii- on-line chronicle recognition', 324–329, (2007).

[19] Thi V. Duong, Hung H. Bui, and S Venkatesh, 'Activity recognition and abnormality detection with the switching hidden semi-markov model', in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, (2005).

[20] D.A. Few, D.J. Bruemmer, and M.C. Walton, 'Improved human-robot teaming through facilitated initiative', in *The 15th IEEE International Symposium on Robot and Human Interactive Communication, 2006. (ROMAN 2006)*, pp. 171–176, (2006).

[21] E. Forman and M. A. Selly, *Decision By Objectives*, World Scientific, 2001.

[22] T. Germa, L. Brèthes, F. Lerasle, and T. Simon, 'Data fusion and eigenface based tracking dedicated to a tour-guide robot', *ICVS*, (2007).

[23] M. Ghallab, 'On chronicles: Representation, on-line recognition and learning', pp. 597–607, (1996).

[24] M. Ghallab, D. Nau, and P. Traverso, *Automated Planning - theory and practice*, Morgan Kaufmann Publishers, 2004.

[25] K. Madhava Krishna, R. Alami, and Simeon T., 'Safe proactive plans and their execution', *Robotics and Autonomous Systems*, **54**(3), 244–255, (March 2006).

[26] Cordier Marie-Odile and Dousson Christophe, 'Alarm driven monitoring based on chronicles', in *In proc. of the 4th Sumposium on Fault Detection Supervision and Safety for Technical Processes (SafeProcess)*, pp. 286–291, Budapest, Hungary, (June 2000).

[27] Luis Marin, Emrah Akin Sisbot, and Rachid Alami, 'Geometric tools for perspective taking for human-robot interaction', in *Mexican International Conference on Artificial Intelligence (MICAI 2008)*, Mexico City, Mexico, (October 2008).

[28] V. Montreuil, A. Clodic, and R. Alami, 'Planning human centered robot activities', in *IEEE Int. SMC*, (2007).

[29] A. J. Schmid, O. Weede, and H. Wörn, 'Proactive robot task selection given a human intention estimate', in *In Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication 2007 (RO-MAN 2007)*, (2007).

[30] J. Schmidhuber, 'Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts', *Connection Science*, **18**(2), 173–187, (2006).

[31] E. Akin Sisbot, Luis F. Marin Urias, Rachid Alami, and Thierry Siméon, 'Spatial reasoning for human-robot interaction', in *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, San Diego, CA, USA, (November 2007).

[32] E.A. Sisbot, L.F. Marin-Urias, R. Alami, and T. Simeon, 'A human aware mobile robot motion planner', *IEEE Transactions on Robotics*, **23**(5), 874–883, (october 2007).

[33] A. Tapus, M.J. Mataric, and B. Scasselati, 'The grand chhallenges in socially assitive robotics', *IEEE Robotics and Automation Magazine Special Issue on Grand Challenges in Robotics*, **14**, 35–42, (2007).

Human to Robot Demonstrations of Routine Home Tasks: Acknowledgment and Response to the Robot's Feedback

Aris Alissandrakis and **Yoshihiro MIYAKE**[1]

Abstract. This paper investigates the possible role of the robot's feedback in Human-Robot Interaction (HRI) from the *human perspective*, and attempts to highlight some important conceptual and practical issues such as the lack of explicitness and consistency on people's demonstration strategies. More specifically, any changes that can be expected on the part of a human (teacher), in the teaching of a task, when a robot (student) declares that the given demonstration was not understood. The findings from such studies can help in turn, from a *system perspective*, towards the design of HRI systems that are able to better anticipate and behave according to human expectations.

Partly intended as a replication and verification of a previous study, the everyday domestic task of setting a table, both in Japanese and in non-Japanese (or "western") style, is taught to a humanoid robot by the participants of the currently conducted user study. The participant's acknowledgment and responses to the robot's feedback are discussed in regard to demonstration changes and consistency, based on a HRI gesture classification.

1 INTRODUCTION AND OVERVIEW

As robots integrate more into human society and domestic/public environments (in contrast to industrial/factory environments), they can be expected to behave more like 'companions', i.e. a) make themselves 'useful' (being able to carry out a variety of tasks in order to assist humans in domestic home environments), and b) behave socially (possess social skills in order to be able to interact with people in a socially acceptable manner). Such robots should take into consideration individual human likes, dislikes and preferences in order to adapt their behavior accordingly.

Thus, issues of agency, believability and sociality become very important [5]. Humans have expectations about the way artifacts should react and perform in social situations and the design of robots needs to conform to some of these expectations. The design of such sociable robots needs input from research concerning social learning and imitation, gesture and natural language communication, emotion and recognition of interaction patterns [6].

Our broader research goal is investigating social learning in HRI and as part of this, how people explain routine home tasks to robots and how these human-robot interactions unfold by using speech and gestures. Using that as a starting point, we want to examine the possible role of feedback by a robot acknowledging (or stating failure to understand) the instructions of a human teacher.

[1] Miyake Lab, Dept. of Computational Intelligence & Systems Science, Tokyo Institute of Technology, Japan, email: alissandrakis@myk.dis.titech.ac.jp

This role would need to be systematically investigated, and examine how if affects the cycle of human-robot interactions. The overall teaching strategy would also need to be informed by studies investigating the characteristic ways of how people tend to segment tasks, forming goals and sub goals explicitly, or implicitly.

We would like to also highlight the dialogic nature of the problem and try to consider two perspectives: (a) *human perspective* – how people do teaching (more-or-less) naturally and (b) *system perspective* – how can hardware and sensor requirements be met and extensive (or even impossible) assumptions/pre-knowledge about the world be addressed. By examining the human perspective, we can better inform the system perspective; this line of research will enable researchers to design robots that can pick "close enough initial metrics of similarity" depending on context, kick-starting the robot's understanding of taught instructions and process of common ground negotiation with the human.

The paper is structured as follows. Section 2 introduces the four research questions. Section 3 details the methodology, section 4 discusses the results from the current user study, while section 5 discusses some design implications from the system perspective. Finally, section 6 offers some conclusions and possible future work directions.

2 RESEARCH QUESTIONS

The issue of how people explain routine tasks to robots has already been addressed in [13], suggesting that people's assumptions about the way a robot "should" be able to understand their demonstration, the implicit knowledge about the tasks that people tend to take for granted, and the robot's behavior legibility can and will influence the HRI in this type of task.

The purpose of the conducted user study is to examine the possible role of feedback (by the robot, to the human) regarding teaching a simple domestic everyday task. This feedback can be either *positive* or *negative*, i.e. indicating understanding (or not) of the human demonstration[2], by the robot.

The work presented here is also intended as a replication and extension of the work presented in [12], and as such shares three main research questions:

Q1 Do the human participants change their instruction(s) when the robot provides negative feedback (i.e. declares inability to understand the participant's gestures and/or speech)?
Q2 What is the nature of the change of the instructions (if any).
Q3 To what extent do people maintain the changes for the remaining of the teaching task.

[2] This can include both verbal explanations, and also non-verbal gestures by the participant.

In addition, given the two variations of the task (Japanese and non-Japanese style of setting a table), we add another:

Q4 Does the nature of the task to be taught (both known, but one not so frequently practiced by the participant) have any influence on the above questions?

On a related note, it is worth mentioning that [14] makes some observations about the way people tend to administer their own feedback when teaching a Reinforcement Learning agent:

(a) they use the reward channel not only for feedback, but also for future-directed guidance;
(b) they have a positive bias to their feedback, possibly using the signal as a motivational channel; and
(c) they change their behavior as they develop a mental model of the robotic learner.

These last points (about people's feedback) are outside the current scope of the work (focusing more on the robot's feedback), but very much of interest in the broad context of this work.

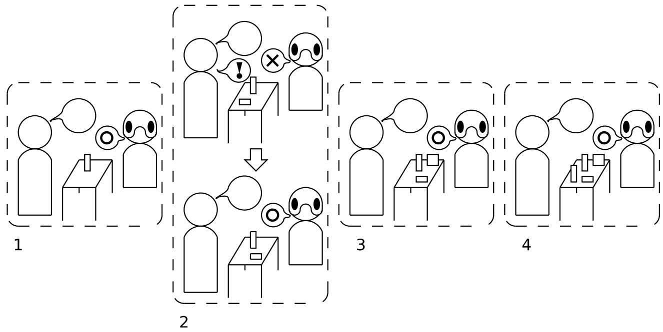


Figure 1. Robot's feedback for each sub-task. Irrespective of what the actual participant demonstration is, the robot should give positive feedback on the first, third and fourth demonstration, and (initially) give negative feedback on the second demonstration. Only after (and if) the participant repeats their second demonstration, the robot should then give positive feedback. See Fig. 2 for the actual gestures used by the robot.

3 METHODOLOGY

3.1 Participants

The sample consisted of 11 native Japanese participants (6 female and 5 male), all of them university students in their early twenties. None of them had any computer science or robotics background, or (more importantly) had any previous experience of interacting with a robot, and could therefore be considered 'naive' in respect to their expectations of a robot behaving in a social manner.

For the purposes of the current exploratory study (as for the previous study presented in [12]) this sample size was considered suitable, in order to provide observations for future studies involving larger number of participants.

3.2 Materials/Apparatus

The current user study was conducted at the Miyake Lab, Tokyo Institute of Technology, in Japan.

Each session consisted of a participant teaching the humanoid communication robot Wakamaru (Mitsubishi Heavy Industries) two variations, Japanese and non-Japanese, of the everyday domestic task of laying a table.

The utensil objects were a plate, a bowl, a cup and chopsticks[3] for the Japanese style, and a plate, a glass, a fork and a knife for the non-Japanese style of laying a table. In each case, they were set on a side-table, from which they should be picked up by the participants and placed on another table, in front of the robot.

A video camera was used to capture the participant's demonstrations and interaction with the robot.

3.3 Procedure

In order to capture the most 'natural' responses and behavior by the participants, no specific instructions were given to them as to how to interact with the robot, besides the single restriction that they must only use one object at the time (but in any order), and hold it using one hand; as there were four objects to be used each time, each task was decomposed to four 'sub-tasks'.

In each session, a participant, after completing a background questionnaire, was asked to demonstrate to the robot how to lay the table twice, using a different set of utensils each time. The order for using the Japanese and the non-Japanese utensils was reversed for each successive participant.

In each of the tasks, the robot would provide positive feedback for the first, third and fourth sub-task demonstrations given by the participant, but (at first) provide negative feedback for the second demonstration (see Fig. 1), irrespective of what the actual instructions by the participants are.

The change from positive to negative feedback between the first and second demonstrations allows us to examine Q1 (and, in the cases that Q1 happens to be true, Q2); the change back to positive feedback for the third and fourth demonstrations allows us to examine Q3, assuming Q1 occurred. By comparing the interaction during the entire task in the Japanese and non-Japanese cases, we can examine Q4.

Finally, each participant was asked in a post-session semi-structured interview to subjectively recall the events as they occurred in their session, and comment on their impression of their interaction with the robot.

3.4 The "Wizard of Oz" Methodology

Due to technological issues (current state-of-the-art robots are not yet able to detect and understand *unrestrained* behavior by humans), plus the fact that for our purposes here the robot does not need to detect or respond to the actual participant's behavior (as all responses are predetermined), the robot can be controlled using the Wizard-of-Oz methodology.

The Wizard-of-Oz has come in common usage in the fields of experimental psychology, human factors, ergonomics, and usability engineering to describe a testing or iterative design methodology wherein an experimenter (the "wizard"), in a laboratory setting, simulates the behavior of a theoretical intelligent computer application or robotic system (often by remaining hidden, intercepting the communication and using teleoperation, with the participant having no a-priori knowledge of this – thus creating the illusion of autonomy). For more details about the Wizard-of-Oz methodology see e.g. [4].

The use of this methodology is very powerful for studies of this

[3] The participants were instructed to consider the two chopsticks together, as a single item. However, one participant (5) also considered the fork and the knife as a single 'item'.

kind; it is important for the current user study as we do not need to do an evaluation from the *system perspective* (therefore the actual robot performance only needs to be consistent and legible throughout) but rather from the *human perspective* observe the participant's behavior to the robot feedback in a controlled way (by having defined predetermined robot reactions for each sub-task).

Although the order and type of the responses is predetermined (as per Figs. 1 and 2), the precise timing depends upon each participant completing their current demonstrations. The end of a demonstration can be defined as when e.g.

- the participant stops talking and/or gesturing,
- the participant starts to interact with the next[4] object,
- the participant waits for the robot's acknowledgment,
- etc.

These criteria are more-or-less subjective, and currently far easier to be detected by humans (sensitive to a variety of cues across different modalities) than by an artificial system; therefore it is important that the wizard is the same person and s/he tries to be consistent throughout the experiment sessions (by using the same criteria in similar situations).

The precise response timing is an issue that cannot be easily controlled in a Wizard-of-Oz study. It has been shown (see [9, 10] for examples in HRI) that besides the explicit verbal part of communication (formal semantics) an implicit non-verbal part (expressed as e.g. the delay between an utterance and its response, but also the coordination of gesture and speech) also plays an important role, both in human-human and human-robot interaction.

3.5 Gesture Classification

The conceptual framework presented in [11] can be used to capture requirements for *contextual interpretation* of body postures and human activities for purposes of HRI. It defines five functional classes of gestures:

Manipulative gestures These are gestures that involve the displacement of objects (e.g. picking a cup), or miming such displacements.

Symbolic gestures These are gestures that follow a conventionalized signal. Their recognition is highly dependent on the context, both current task and cultural milieu (e.g. the thumbs up or thumb-index finger ring to convey "OK").

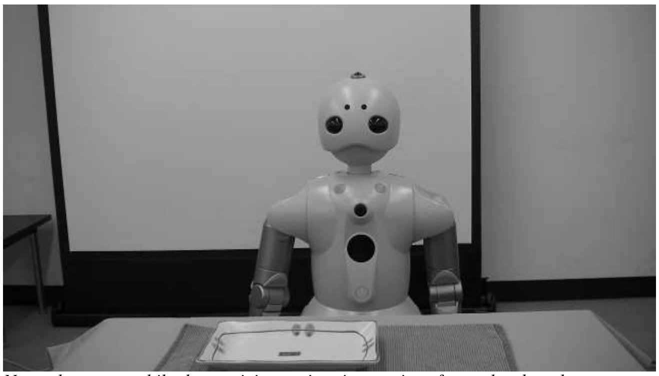
Interactional gestures This category classifies gestures used to regulate interaction with a partner. These can be used to initiate, maintain, invite, synchronize, organize or terminate an interaction behavior between agents (e.g. head nodding, hand gestures to encourage the communicator to continue).

Referencing/pointing gestures (Deictics) The gestures that fall into this category are gestures used to indicate objects or loci of interest.

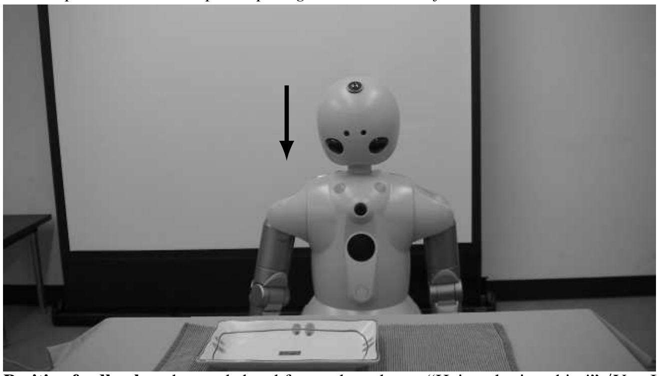
Side effect of expressive behavior These are gestures that occur as side-effects of people's communicative behavior. They can be motion with hands, arms, face etc but without specific interactive, communicative, symbolic or referential roles.

Irrelevant These are gestures that do not have a primary communicative or interactive function, e.g. adjusting one's hair or rubbing the eye.

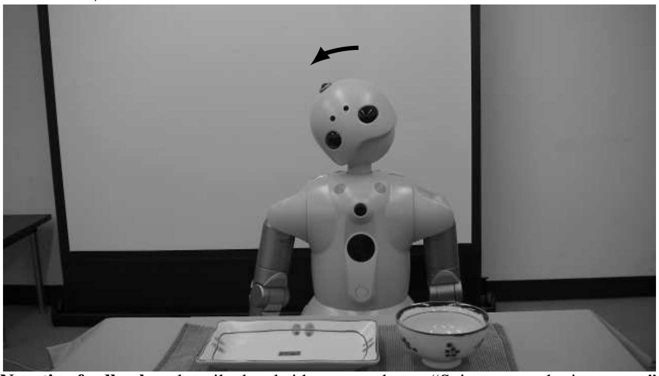
[4] Note that transporting and placing an object is usually not the end of a demonstration, and should not be used as such in general.



Normal posture, while the participant gives instructions for each sub-task.



Positive feedback: robot nods head forwards and says "Hai, wakarimashita!" ⟨Yes, I understand!⟩



Negative feedback: robot tilts head sideways and says "Suimasen, wakarimasen..." ⟨Sorry, I don't understand...⟩

Figure 2. Robot feedback. From top to bottom: normal posture, positive and negative feedback.

But it is important to note that certain gestures in particular situations might be multipurpose. Nehaniv et al. [11] stress the importance of knowing the context in which gestures are produced since it is crucial to disambiguate their meaning. In practice, data on the interaction history and context may help the classification process. The above framework is intended to complement existing and more detailed speech-focused classification systems (see for example [2, 3, 7]).

Towards examining the research questions posed in section 2, one can analyze the video obtained from the user study sessions, using the above classification. Doing this, a qualitative and descriptive picture of the interactions that took place in the sessions can be obtained, which can be also used to inform the requirements, from the system perspective, for identifying the human instructions and responding accordingly.

For the task involved in this study, similar as in [12], only three of the previous categories were hypothesized to be relevant: *interactional gestures*, *manipulative gestures* and *referencing/pointing gestures*.

As utterances provide contextual information, one can also consider the following classification:

Interactional utterances produced to initiate, maintain, regulate or terminate the social part of the interaction.

Non-interactional utterances produced for the actual explanation of the proposed task. They can be further sub-divided into *indication of object*, *action* and/or *location*.

The objects' localization type can be also classified as

Relative reference, when the person indicates other objects or landmarks as reference points to the location where a specific object should be laid.

Absolute reference, when the person does not use other objects or landmarks as reference points. For example, the person might say "place the glass here" without making any additional gestures, apart from the actual transportation, to indicate possible relations among objects or landmarks regarding location.

Note that this localization classification is related (from the system perspective) to *effect metrics* for robot imitation (cf. [1]) i.e. measures of similarity between positions, orientations and states of external objects, but also changes to the body-world relationship of the agent, that can be used to achieve the same goal(s) in context.

4 RESULTS

The given task had no 'right' or 'wrong' way to be accomplished. The layout of the eating utensils on the table was done according to the personal preferences of each participant, with no objective measure of the task performance. Instead, this section presents a descriptive analysis of the data obtained from the user study; these include the recorded video from the sessions as well as the post-session interviews with each the participants.

As the current user study is (intentionally) very similar to the one presented in [12], we will also contrast some of the respective findings; that study will be referred to in the text as "previous", while the presently reported study as "current".

One general observation that is very contrasting to the previous study was the almost total lack of explicit referencing/pointing gestures. Only a single participant (3, in Japanese utensil task) used some explicit pointing gestures to indicate (relative) locations, but not to indicate an object. Four participants (4,7, 8 and 10, in both tasks) used a single, continuous, gesture to transport each object from the side table to the table in front of the robot, while at the same time verbally identified the object – sometimes adding localization information (4 and 10 only). The rest of the participants (1, 2, 5, 6, 9, 11 and including 3) used an implicit type of pointing gesture, typical, among other places, in Japan for presentation: the object was held at the end of the extended arm[5] of the participant, towards the robot, for a short[6] time before placing it on the table. This kind of gesture could be interpreted as both referencing the object itself but also its intended destination, although for the later interpretation, the extended arm's direction would not give precise information. Therefore, similar to the previous study, there was a lack of clear gesture segmentation between identifying an object and indicating its intended placement.

In some cases there were some subsequent correction gestures, moving the other objects on the table to improve the available space or to better orient the current utensil, but with no added verbal explanation as to what they were doing or why. Overall, there were no distinct interactional gestures observed. Two of the participants (4 and 6) produced an excessive amount of irrelevant type gestures throughout their sessions. Also, although most participants faced the robot across the table, participants 8 and 10 chose instead (for both tasks) to approach the side of the table, and lay the table while standing on the right side of Wakamaru.

[5] Some participants, contrary to the user study instructions, used both hands.

[6] Only one participant (6, both tasks) waited long enough after he presented a utensil in this manner until he explicitly received acknowledgment that the robot perceived the object, before demonstrating where the object should be placed.

4.1 Acknowledgment of the Robot's Feedback

In respect to Q1, every participant acknowledged the robot's feedback, and repeated (modified in some respect) their demonstrations for the second sub-task, in both tasks, when the robot stated that it does not understand their initial instructions.

4.2 Modification of Sub-Task Demonstration

The majority of the participants (1, 2, 4, 8, 9 , 10 and 11) when asked about it in their interviews, stated that the misunderstanding during the second sub-task (for either one, or both of the tasks) occurred due to their own fault (e.g. they did not speak loud enough for the robot to hear, or they did not hold the object within the robot's point of view); therefore, they repeated their demonstration more-or-less as before, both in terms of gesture and utterance, but speaking in a more loud and clear manner.

In terms of object localization, participants used both absolute (e.g. "I place it *at the center*.", "I place the plate *here*.") and relative (e.g. "I place the glass *on the right from your view*.", "I place the fork *on the opposite side*.", "I place the fork vertically *on Wakamaru's right*, *next to* the knife.") referencing, although in some cases they simply omitted any spoken placement instructions. Table 1 shows the object localization classification results for the current user study, indicating only three cases where the type of reference changed, as a response to the robot's misunderstanding. Note that for classification purposes, we considered the occurrence of an *absolute* reference when the participant explicitly used either "here" or "center", with no additional points of reference. In that sense, the absence of an explicit utterance but instead the performance of an implicit (*manipulative* or *deictic*) gesture could be also classified as an instance of absolute reference. However, in Table 1 we differentiate between absolute and absence-of (explicit) localization ("none").

In terms of instruction detail, all seven participants that used relative localization (1, 2, 3, 4, 6, 9 and 10) modified their instructions by changing or clarifying the reference points used (e.g. "[. . .] place it on the right back." → "[. . .] a little more in the back than the plate, to the right as much as it can be." or "I place it on the right." → "I place it on the right of Wakamaru."); three of those participants (1, 6 and 9) stated in their interviews that they thought the misunderstanding for the second sub-task occurred because they did not explicitly specify whether by "left" or "right" they meant from their own or the robot's perspective. In general, the instruction detail increased, with only a single participant (6) choosing to use 'simpler' referencing ("[. . .] on

#	task type	sub-task 1^{st}	2^{nd}	3^{rd}	4^{th}
1a	J	absolute	relative	relative	relative
1b	nJ	absolute	relative	relative	relative
2a	nJ	absolute	relative	relative	relative
2b	J	relative	absolute→relative	relative	relative
3a	J	relative	relative	relative	relative
3b	nJ	relative	relative	absolute	relative
4a	nJ	absolute	relative	none	none
4b	J	relative	relative	relative	absolute
5a	J	absolute	none	none	none
5b	nJ	none	absolute	relative	–
6a	nJ	absolute	relative	absolute	relative
6b	J	absolute	relative→absolute	absolute	absolute
7a	J	none	none	none	none
7b	nJ	none	none	none	none
8a	nJ	absolute	absolute	none	none
8b	J	absolute	absolute	none	none
9a	J	relative	relative	relative	relative
9b	nJ	absolute	relative	relative	relative
10a	nJ	absolute	relative	relative	relative
10b	J	relative	absolute→relative	relative	relative
11a	J	absolute	absolute	absolute	absolute
11b	nJ	absolute	absolute	absolute	absolute

Table 1. Object localization classification results (based on each participant's utterance). Task type order as given to each participant (*J*=Japanese, *nJ*=non-Japanese utensils). *None* indicates no explicit reference (i.e. only utensil name used). Only in 2b, 6b and 10b the participants changed the localization for the second sub-task. In all other cases, it remained the same. In task 5b, participant 5 incorrectly used both knife and fork together as the first item, similar to the chopsticks.

this side." → "[. . .] here." and "[. . .] on the right from your view." → "[. . .] on this side.")

One participant (7), when faced with the robot's misunderstanding in the second sub-task, she placed the utensil back on the side table, chose a different one instead, and then later came back to that utensil for the next sub-task. This occurred for both tasks, although there was no mention that the presentation order was important in the pre-session instructions given to the participants.

4.3 Consistency of Task Demonstration

As seen in Table 1, in every case when a participant changed their object localization reference type (2b, 6b and 10b), they remained consistent through the remaining sub-tasks; this is in contrast to the observations of the previous user study, where only few participants did. However, in the majority of cases in the current study, the participants did not change their referencing at the second subtask (in contrast to the previous study, where a larger number did), and besides participants 7 and 11 (which used "none" and absolute referencing, respectively, for the entirety of both tasks) a combination of localization types was used.

Although it could be argued that most participants in the current study were more-or-less consistent on their own, it is evident that there is still no consistency between them; even a simple assumption like e.g. "absolute reference is most likely to be used for the first object, in the absence of task-related reference points", cannot be generalized.

In comparison to the previous study, the current study results seem to indicate a higher degree of consistency, especially in the cases when a change was made in response to the robot's feedback; however the observation that in general humans lack consistency in HRI teaching tasks remains an issue, especially when (as in these initial user studies) the robot's feedback is not more informative besides a simple indication of understanding (or not).

4.4 Influence of Task Familiarity

In respect to Q4, in the current study, no differences were observed in the demonstration either of the whole task or in specific sub-tasks, between the two tasks, in terms of gestures. However, four participants (2, 3, 5 and 9) commented on the particular function of the Japanese utensils (e.g. "This is a *soup* bowl [. . .] it is used to eat miso soup.", "This is a *rice* bowl. I use it when I eat rice.")[7] compared to simply labeling the non-Japanese utensils. None of the participants added this description detail to their instructions for any of the non-Japanese utensils.

5 SOME STRATEGIES FROM A SYSTEM PERSPECTIVE

For currently available algorithmic and machine learning methods, the demonstrations by the participants of this study were far from clear or easy to identify. The setting of the study was quite freeform – the participants were given minimal instructions. The system/robot's capabilities were purposefully underpublicized (and we deliberately used 'naive' participants), in order to elicit a wide range of 'natural' responses from the participants, especially when the robot states that it can not understand their demonstration. We believe that investigating what people say and do while interacting with (here demonstrating to) robots, as well as what the people think about the robot's understanding is quite important for HRI. This line of research can enable researchers to design robots that are able to engage in a process of common ground negotiation with humans, depending on the context, towards achieving a variety of tasks/goals.

From a human perspective, we note that people are willing to interact with robots, but most importantly that they appear willing to accommodate the robot's feedback and modify their teaching demonstrations to facilitate its understanding. However, consistency of teaching style aside, without knowing what was unclear about the initial example, the additional example might contradict task-specific knowledge indicated in the previous example – or again be unclear. Given the initial character of the current study, the feedback was rather simple – an acknowledgment of understanding (or not), but not *what* was understood. A system should be able to communicate the nature of the source of the misunderstanding; one way this could be achieved would be for the robot to (partially) reproduce the current sub-goal. Even without verbal comments from the robot, such a reproduction attempt would implicitly advertize the robot's capabilities, and could provide the human with an insight on the source of the failure, to be addressed in their next demonstration attempt. This 'interactive', turn-taking approach to teaching would allow robots to take full advantage of the social environment for learning, and humans to increase their communication potential.

One of the reasons for this misunderstanding might be that people tend to combine information about the object and the manipulation/transportation/location. Unprompted, there is usual no clear segmentation, either by a distinct pause or particular gestures that emphasize e.g. the object or the location. From a system's perspective then, it is important for the robot to be able to prompt people to

[7] Note that both participants are here referring to the same utensil.

be more explicit about the sub-task segmentation of their demonstrations. This would possibly help isolate and identify any sources of misunderstanding, and target them specifically.

6 CONCLUSIONS AND FUTURE WORK

Based on the results from the current user study, all[8] participants did acknowledge the robot's feedback on their demonstration of teaching the task of laying a table (Q1); they appreciated the positive feedback, but most importantly they responded to the negative feedback as well, by repeating their demonstration. Unfortunately, the participants used very few distinct gestures, and, as the negative feedback was not very informative, the majority assumed that e.g. they had to speak louder, so they did not greatly modify their demonstrations (Q2). However, in the few cases they did (change the object localization), they were consistent for the rest of the task (Q3) – this observation was more promising than the findings of the previous study. Concerning the influence of task familiarity on the current task teaching (Q4), there were no definitive results from the current study, but it would appear that the participants tended to add extra details about the object's intended use and function in their verbal instructions.

One of the main motivations for the current user study was to confirm whether the initial observations of the previous study [12] on the acknowledgment of the robot's feedback were valid in a Japanese cultural setting (also, using a humanoid[9] robot), and inform the follow-up studies.

Next research topics would include further studies on to what extent are people willing to engage in interactions with robots that actively seek to show their understanding by demonstrating, and whether this strategy would be realistic in a wide range of domestic situations. Given the limitations of state-of-the-art active manipulation, what other mechanisms could perhaps put into place and serve as replacements?

Are people willing to accept the robot's requests regarding the way a task can be performed? What form should these requests take? Regarding these last issues, in the follow-up study we intend to increase the details of the robot's feedback, by adding details about the object's localization, followed by a pointing gesture. This could be either absolute ("here") or relative (e.g. " to the left of [name of other utencil]"), and instead of positive or negative, the feedback could be 'correct' (according to the participant's perceived intentions by the 'wizard') or not. A possible strategy to explore would then be to have Wakamaru use only one of either localization styles and make 'mistakes' if the participant uses a different one (or 'none').

Another observation we made on the current study is that most participants appeared during their session (and some made a remark in their post-session interviews) to expect more interactional gestures on the part of Wakamaru, instead of the robot maintaining a passive non-moving posture while they were performing their demonstrations. Therefore, we also aim to have Wakamaru actively indicate that it is being attentive (e.g. by regularly uttering "Hai ⟨Yes⟩" and nodding), while the sub-task demonstration takes place, before the feedback – this sort of indication of attention is very common (and expected) in Japanese culture.

One participant also remarked in her interview that the robot having the same response time for both positive and negative feedback did not feel right; in her opinion, the expression of misunderstanding should be more delayed (indicating perhaps a longer and more careful thinking process, or even politeness). In a follow-up system, more autonomous and less Wizard-of-Oz, the response timing of the feedback (time between the end of the participant's utterance and beginning of the robot's utterance) should be controlled as a function of the duration of the participant's utterance (similar to the approach in [9, 10]). This timing synchronization would facilitate a "co-creation" process [8], meaning the co-emergence of real-time coordination by sharing subjective time and space between different persons (and robots).

[8] But this is not always true; e.g. in the previous study a single participant chose to completely ignore the robot's feedback. More study is needed regarding this hypothesis in general.

[9] In the previous study, a robot of mechanistic appearance (PeopleBot, MobileRobots Inc.) was used.

ACKNOWLEDGEMENTS

The work presented in this paper was supported by a FY2007 (P07369) Postdoctoral Fellowship for Foreign Researchers from the Japan Society for Promotion of Science (JSPS).

The first author would like to acknowledge the hosting by Prof. Yoshihiro MIYAKE in Tokyo Institute of Technology. Both authors would like to thank Shohei YOSHIDA, Tatsunori NISHI and Kengo OKITSU for help in conducting the user study, Yukiko Thuresson for transcription and translation of the interviews, and Yumiko MUTO for additional translation and help.

REFERENCES

[1] A. Alissandrakis, C. L. Nehaniv, and K. Dautenhahn, 'Action, state and effect metrics for robot imitation', *Robot and Human interactive Communication, 2007. RO-MAN 2006. The 15th IEEE International Symposium on*, 232–237, (Sept. 2006).

[2] Justine Cassell, 'Nudge nudge wink wink: elements of face-to-face conversation for embodied conversational agents', in *Embodied conversational agents*, ed., J. Cassell et al., 1–27, MIT Press, Cambridge, MA, USA, (2000).

[3] Justine Cassell, Stefan Kopp, Paul Tepper, Kim Ferrimanand, and K. Striegnitz, 'Trading spaces: How humans and humanoids use speech and gesture to give directions', in *Conversational Informatics*, ed., T. Nishida, John Wiley & Sons, New York, (2007).

[4] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg, 'Wizard of oz studies—why and how', in *Readings in intelligent user interfaces*, 610–619, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, (1998).

[5] Kerstin Dautenhahn, 'The art of designing socially intelligent agents: science, fiction and the human in the loop', *Applied Artificial Intelligence Journal, Special Issue on Socially Intelligent Agents*, **12**(7-8), 573–617, (1998).

[6] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn, 'A survey of socially interactive robots', *Robotics and Autonomous Systems*, **42**(3-4), 143 – 166, (2003).

[7] D. McNeill, *Hand and Mind*, University of Chicago Press, Chicago, 1992.

[8] Y. Miyake, 'Co-creation in man-machine interaction', *Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on*, 321–324, (2003).

[9] Y. Muto, S. Takasugi, T. Yamamoto, and Y. Miyake, 'Timing control of utterance and gesture in human-robot interaction'. submitted, 2009.

[10] K. Namera, S. Takasugi, K. Takano, T. Yamamoto, and Y. Miyake, 'Timing control of utterance and body motion in human-robot interaction', *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, 119–123, (Aug. 2008).

[11] C. Nehaniv, K. Dautenhahn, J. Kubacki, M. Haegele, C. Parlitz, and R. Alami, 'A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction', in *Proc. IEEE Ro-Man, 2005*, pp. 371–377, (2005).

[12] Nuno Otero, Aris Alissandrakis, Kerstin Dautenhahn, Chrystopher Nehaniv, Dag Syrdal, and Kheng Lee Koay, 'Human to robot demonstrations of routine home tasks: Exploring the role of the robot's feedback', in *Proc. 3rd ACM/IEEE International Conference on Human-Robot*

Interaction (HRI2008), Amsterdam, Holland, March 12-15, 2008, pp. 177–184, (2008).

[13] J. Saunders, N. Otero, and C.L. Nehaniv, 'Issues in human/robot task structuring and teaching', *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, 708–713, (Aug. 2007).

[14] Andrea L. Thomaz and Cynthia Breazeal, 'Teachable robots: Understanding human teaching behavior to build more effective robot learners', *Artif. Intell.*, **172**(6-7), 716–737, (2008).

Supporting Remote Manipulation with an Ecological Augmented Virtuality Interface

J. Alan Atherton and **Michael A. Goodrich** [1]

Abstract.
User interfaces for remote robotic manipulation widely lack sufficient support for situation awareness and, consequently, can induce high mental workload. With poor situation awareness, operators may fail to notice task-relevant features in the environment often leading the robot to collide with the environment. With high workload, operators may not perform well over long periods of time and may feel stressed. We present an ecological visualization that improves operator situation awareness. Our user study shows that operators using the ecological interface collided with the environment on average half as many times compared with a typical interface even with a poorly calibrated 3D sensor; however, users performed more quickly with the typical interface possibly because of the poor calibration.

1 INTRODUCTION

A manipulator is a machine that can operate on its surrounding environment. Manipulators are commonly modeled after a human arm and hand, and aim to perform actions similar to what humans can do with their arms and hands. A *remote* manipulator is located away from the operator, so that the operator depends on sensors near the robot to provide information about the remote environment. Figure 1 shows an example of a remote manipulator and supporting sensors. A *mobile* manipulator is manipulator mounted to a mobile base, and is often also a remote manipulator.

Mobile manipulators are useful tools in areas that are dangerous or inaccessible to humans. For example, they are used in planetary exploration (Mars rovers), urban search and rescue (USAR), and explosive ordnance disposal (EOD) [16, 2, 17]. Operators use the manipulator to grab, push, poke, operate tools, and generally try to do what people do with their hands. In order to safely benefit from the capabilities these robots provide, a human operator must control them remotely. Operators can use the mobile manipulator to perform a variety of tasks from a safe location. Although supporting mobile manipulation is our end goal, the complexity involved is more than we can test in one study, so we limit the scope of this work to remote manipulation as a step toward supporting mobile manipulation.

One difficulty in remote manipulation is for the operator to understand the situation given limited data. In remote situations, the only connection the operator has with the robot is the user interface, and the operator is limited to whatever information the robot or the interface can provide. This is sometimes referred to as looking at the world through a "soda straw" [23]. Problems that come from this limited view of the world include missed events, difficult navigation in new situations, and an incomplete understanding of the explored world [23]. For example, depth perception is a very important aspect of understanding the environment for manipulation, but single-camera video displays provide limited depth perception in unstructured environments.

[1] Department of Computer Science, Brigham Young University, USA

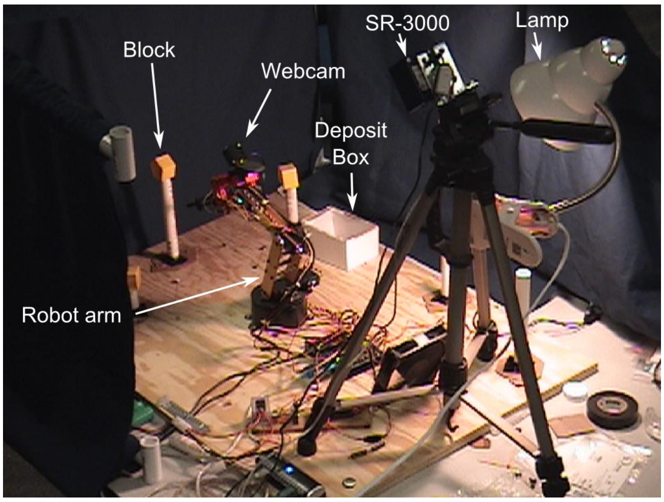


Figure 1. Remote manipulator robot, surrounding environment, and supporting sensors.

Factors that affect performance in remote manipulation tasks include the physical capabilities of the robot, the autonomous behaviors of the robot, the user interface, and the capabilities and skill of the human operator. Although all of these factors affect performance, we can focus on a few factors that a system designer can control. Assuming that we have a skilled operator, the design variables that we can freely manipulate include the following key elements: robot capabilities, the robot behaviors, and the user interface. When the robot does not have the required capabilities for a task, then the usability of the user interface or autonomous behaviors make little difference for accomplishing the task. On the other hand, when the robot does have the necessary capabilities, then the user interface and autonomous behaviors can significantly affect performance.

Typical robotic user interfaces present several sets of information in multiple windows with each window showing a distinct set of information (see, for example, Figure 2). When information is displayed disjointly in this manner, operators must mentally integrate the displays to understand relationships between different sets. While such interfaces can be very useful, including for testing and debugging a system, such a design can increase the mental workload on operators. By splitting the information into multiple windows, operators have poorer manipulation or task-specific situation awareness as they may miss an important event from one display while focused on a different display.

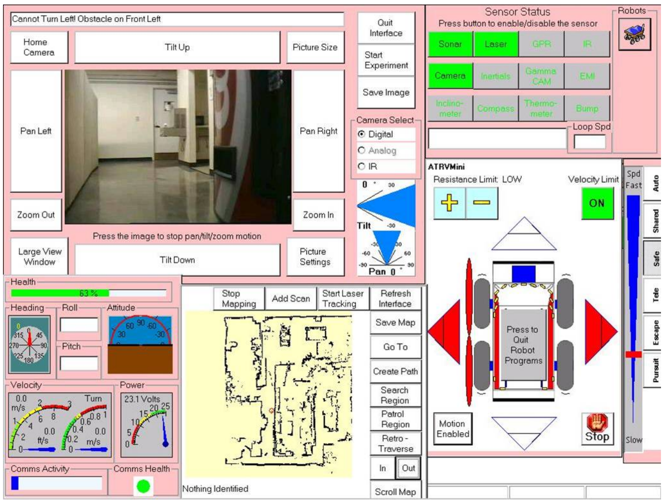


Figure 2. Idaho National Laboratory (INL) mobile robot control interface with multiple windows. Adopted from [1].

We present an ecological augmented virtuality (AV) user interface. As discussed by Vicente in [20], the term *ecological interface* originates from psychologist James J. Gibson who defines *ecological psychology* as study of human-environment relationships in a natural setting, as opposed to a laboratory setting. Ecological interface design aims to make relationships in the environment perceptually evident to the user in order to minimize workload for understanding those relationships. The term *augmented virtuality*, as defined by Paul Milgram, means the interface presents real-world information (e.g. camera video) inside a virtual environment [11]. Augmented *reality*, on the other hand, displays virtual information on top of camera video.

The ecological AV interface is inspired by work done by Milgram [12], Ricks [15], Nielsen [14], and Michaud [10]. The primary component of the interface is an ecological visualization for obstacles and objects around the robot. The visualization displays a 3D range scan in context with a virtual robot arm to show the spatial relationship between the robot and its surroundings; see Figure 3. We claim that the visualization gives the operator more viewpoints than are typically afforded by video camera systems and provides a grounded frame of reference, potentially improving performance.

The primary contribution of this research is the analysis of the benefits of using an ecological AV interface to support robotic manipulation. Another contribution is the interface design itself; in particular, the ecological 3D range scan visualization. We examine the effects of such an approach on overall performance and situation awareness.

For the remainder of the paper we will show how our approach improves situation awareness in a remote manipulation task. We discuss the motivating factors in the design of our user interface. We describe our experiment design, explain how the results show improved situation awareness, and present conclusions.

2 RELATED LITERATURE

Milgram and Drascic et al. show that stereo vision displays improve performance and accuracy in manipulation tasks, and that virtual tools augmenting the display further improve performance and accuracy [12]. People were able to more accurately position a robot arm with virtual tools including tethers, tape measures, pointers, landmarks, and object overlays. In this paper we look at a method that gives some of the benefits of stereo vision without the high bandwidth requirements and viewing hardware.

Nguyen et al. present several NASA virtual reality interfaces for scientific and space robotics, the most similar to ours called Viz [13]. Viz is a highly flexible, modular interface system that can display virtual models of robots in context with data from various sources, such as automatically generated 3D terrain models. The interface and visualization we present is similar in many ways to Viz; however, we have designed with real-time operation in mind, while Viz seems to be designed more for planning due to the long delays associated with extraplanetary robotics.

Nielsen shows that an augmented virtuality interface improves performance for mobile robot navigation and exploration tasks [14]. The interface displays a live video feed from a pan-tilt-zoom camera in context with a 2D map built with a laser rangefinder. This paper explores whether there are similar improvements when applying ecological augmented virtuality design to a manipulation task.

Tsui and Yanco et al. designed an interface for tasking a wheelchair-mounted manipulator arm [19]. Unlike other related work, the operator is in close proximity with the robot, so there is less need for situation awareness. The interface facilitates interaction between the robot arm and the operator. The operator indicates an object of interest with a joystick or touchscreen, then the robot arm autonomously moves close to the object.

3 METHODS

Our method of improving user interfaces for remote manipulation has four steps: choose a relevant problem, identify interface goals and requirements for the problem, design interfaces using different approaches, then test to see which approach is better. Thus, we break up the discussion of methods into four respective sections.

3.1 Application

The real-world application most similar to what we will use for experimentation is remote sample acquisition. The objective in sample acquisition is to navigate the robot through an environment and use a manipulator to collect samples, such as rocks, and analyze them onsite or return them to a lab for further analysis [16]. We will mimic this task and analyze the performance of operators using a variety of interface designs.

3.2 Interface Goals and Requirements

To guide the design of our interface, we choose some goals and requirements inspired by Goodrich and Olsen [5]. The primary goals are to:

1. maintain a manageable workload and
2. support situation awareness.

Requirements for the interface are to:

1. integrate information from multiple data sources,
2. provide views of the environment from multiple perspectives,
3. act as a grounding frame of reference for interaction with the robot, and
4. externalize memory.

3.2.1 Goals

Maintain a manageable workload.

A manageable workload means that the operator can perform the tasks without feeling overly stressed, operate in a situation where there are multiple competing tasks, and sustain operation for some period of time [18, p. 301]. Reducing the operator workload is especially important in EOD and USAR, where operators may have to work while physically and mentally fatigued [2]. To reduce the workload, we must (a) think about information presentation that leverages human perception and (b) devise control methods so that users can build correct, simple mental models [21, p. 132].

Support situation awareness. Endsley defines situation awareness as "The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [4]. In the context of remote manipulation, situation awareness involves understanding the environment surrounding the robot to avoid obstacles and to position the manipulator with sufficient precision to accomplish a task.

3.2.2 Requirements

Integrate information. When information coming from different sources is presented in separate contexts, it can be difficult to see and understand relationships that may exist between different types of data. For example, if robot position is displayed as text on one computer screen, and a map of the environment around the robot is displayed as an image on a separate computer screen, it may be difficult to quickly determine where the robot is located on the map. Integrating information can reduce mental workload by reducing the number of mental transformations that must take place to establish relationships between different sets of data [22, p. 189].

View the environment from multiple perspectives. Providing the ability to view multiple perspectives can help operators to acquire situation awareness, especially since they must understand a 3D environment on a 2D display. Several views can help the operator understand the relationship between the robot and its surroundings. This is particularly useful in manipulation tasks where depth perception plays a large role. Although a single camera only provides a single perspective, creating a virtual environment can be a way to provide multiple perspectives [14, 3]. With a virtual environment, the operator can move a virtual camera around to change perspectives.

Grounded frame of reference. Macedo et al. [9] describe how people perform better when information display is grounded with control. This means that the robot moves in a way the operator expects. When the display is rotated it may not immediately be clear what direction the robot arm should move when given a command. Grounding simplifies the operator's mental model of how the controls affect the robot arm by connecting the motion of the arm to the display [21, p. 135]. A simpler mental model means that the mental workload is lower and performance is higher. More recent work by Hiatt et al. suggests that a task-centric frame of reference yields better performance than robot-centric or view-centric frames of reference [6].

Externalize memory. When the interface includes real-time data, such as robot telemetry or a live video stream, it can be difficult to remember what happened in the past. For example, if an obstacle comes into view on the camera, then leaves the view, it is difficult to remember where the obstacle is after a while [5]. Relieving burden on short-term memory can help reduce workload and improve performance. Externalizing memory means we place information in the interface so the operator no longer needs to keep what happened in working memory [22, p. 191].

3.3 Interface Design

We designed a new interface with these requirements in mind. The interface is a type of 3D ecological AV display similar to the work done in [12, 14, 15, 10]. Our interface differs from [12] in that they use stereo vision with a fixed camera viewpoint and overlay virtual elements as in traditional augmented reality, whereas we create a virtual environment to display a 3D model that can be seen from several viewpoints. While [14, 15, 10] apply ecological AV interfaces to real-time robot navigation, we look at this type of interface for real-time robotic manipulation.

In order to reduce workload on the operator and improve situation awareness, we implemented a display that: (1) integrates information from different sources, (2) provides views from multiple perspectives, (3) gives a grounded frame of reference, and (4) externalizes memory. We explain the design of each of these parts and how each meets the requirements in the previous section.

Integrating information into one display may reduce the number of mental transformations the operator must perform to establish the relationships between different data sources, such as robot state, map, manipulator arm state, and camera video. We integrate information by rendering virtual representations of the data sources such that their spatial relationships are directly perceivable. We render a virtual graphic of the manipulator arm that reflects the position and orientation of all the robot's parts. We also display a 3D scan generated from the range image data that we explain in greater detail later in this section. See Figure 3 for a screen-shot that depicts these representations, and also shows the view the user has while using the interface. By integrating these sources into a single display, the mental workload on the operator is potentially reduced [15].

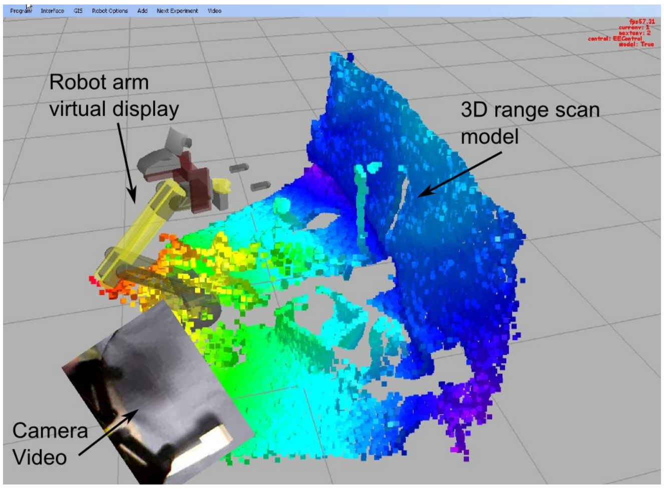


Figure 3. Augmented virtuality user interface for remote manipulation. This is the user's view in the interface.

Since we display information in a virtual environment, we can provide **virtual viewpoints** from anywhere in the environment. The user is enabled to change the "look-from" point of the virtual camera, and the focal point of the camera is set on the center of the workspace of the robot. These adjustments give multiple perspectives to the user and enhance understanding of the world surrounding the robot arm.

While multiple perspectives are nice, they introduce a problem with controlling the robot arm. When the user indicates a direction for the robot arm to move, should the robot move according to its own frame of reference or from the virtual perspective's frame of reference? In essence, this is a problem of **grounding**. We implemented two control methods: the *robot frame-of-reference control* method is direct, independent control of each joint, and the *grounded control* method is a "flying the end effector" control method where the user controls an invisible target point to which the arm reaches. In other words, the user moves the target position for the gripper, and the robot decides how to move each joint to reach that position. Controlling the end effector has been shown to reduce workload dramatically compared to controlling the joints, although situation awareness is negatively impacted when controlling the end effector [7]. We hypothesize that the negative situation awareness effects will be reduced as a result of the integrated nature of our display. With end-effector control, the orientation of the virtual view affects the direction the target point moves (see Figure 4). In other words, commanding the target point to move right will cause the target to move right in the computer screen frame of reference, and not necessarily in the robot's frame of reference. This can cause confusion while focused on the camera video (because the video is in the robot's frame of reference), so we rotate the camera video to give a sense for what direction the robot will move relative to the camera video. The grounded control connects the control of the robot arm to the virtual display of the world.

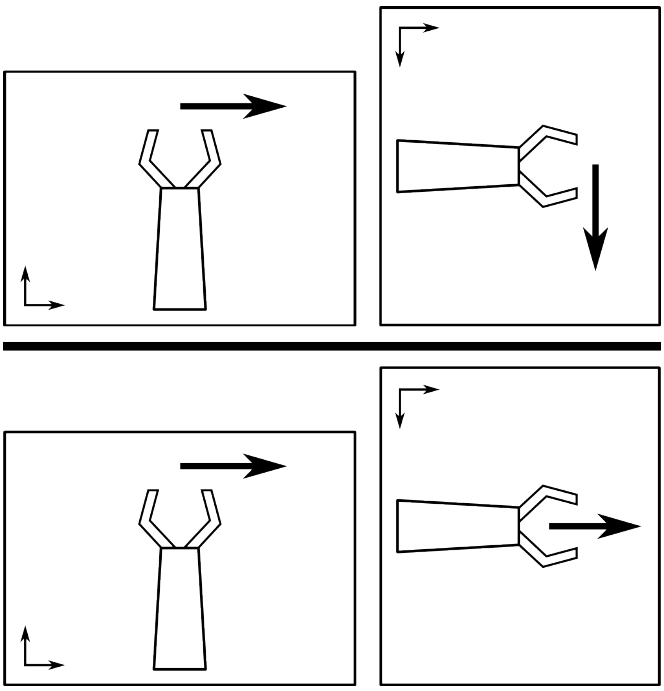


Figure 4. Two frames of reference for controlling a robot arm. All boxes show the robot arm from a top-down perspective. The robot arm is in the same position in each box, but the view is rotated. The top row shows the direction the end effector would move when commanded to move right when using a robot-centric reference frame. The bottom row shows the direction when using a view-centric reference frame.

Part of the world around the robot is represented by a 3D scan. The scan is captured from a ranging camera (but only when the user requests the scan) and then displayed as a set of colored splats. The color assigned to each splat is a function of depth in a "heat map" gradient. We align the 3D scan with the graphic of the robot arm simply by manual measurements that result in a degree of misalignment. The misalignment is not constant and in some cases appears to be perfect, but in some cases is off by as much as 2 cm. We leave improved alignment to future work.

The 3D scan presents a perceptual display of information about the world and thus potentially improves situation awareness while reducing workload. Such a display also provides **externalized memory**, as the 3D scan provides a greater field-of-view than the video camera.

4 User Study

In the user study we evaluate the usefulness of the 3D scan compared with the camera video as well as two different control methods: (1) joint control and (2) end-effector control. We hypothesized that users would perform tasks most quickly and with the fewest collisions while using the 3D scan, camera video, and end-effector control together. We measure the performance of users as they collect blocks with a robot arm in two control-type conditions (joint and end-effector) and three visualization conditions (scan-only, video-only, and scan+video).

4.1 Experiment Setup

For our experiment, we designed a simple object collection task. Participants in the experiment were required to pick up 3 small yellow blocks (2.8 cm wide) and drop them into a cardboard box (8W x 10L x 6H cm). These target blocks are placed on top of posts with heights 6, 10, and 18 cm. Seven different layouts specify the position for each post and the deposit box to prevent users from memorizing the course. In most of the layouts, the posts are placed within relatively easy reach of the robot arm, but a few configurations require the operator to almost fully extend the arm to reach the block. Curtains obscure the robot from the operator's view, although audio cues are present (primarily when a block falls from its post or is deposited).

The robot system is comprised of a robotic arm, sensors, computers, and a joystick controller. See Figure 1 for a photograph of the experiment setup.

The robot arm used in the experiment is a modified Lynxmotion 6 degrees-of-freedom (DOF) arm (5 DOF + gripper). The modified arm can reach approximately 30 cm and lift approximately 0.5 kg. The gripper measures 5 cm when fully open, so there is a 2 cm clearance between the gripper "fingers" and the target blocks. The robot servos provide no position or force feedback. We modified the arm with higher-power servos and added position feedback for the gripper. Feedback on the gripper allows us to display the gripper correctly when an object is grasped; otherwise, the gripper would appear completely closed when grasping an object even though it should appear only partially closed.

The experiment used a ranging image sensor and a webcam. The ranging image sensor is a Swiss Ranger SR-3000 mounted on a stationary tripod above and behind the robot arm. This positioning simulates a configuration where the sensor is mounted above and behind the arm on a mobile robot base. In a real mobile manipulator situation, the base is often stationary during manipulation. Future work will consider the arm mounted to a mobile base. Upon user request, the SR-3000 provides a 3D point for each pixel in the image (dimension 176 by 144 pixels). A generic USB color webcam, mounted to the arm, has a resolution of 320 by 240 pixels and a frame rate of approximately 8 frames per second in our environment. There is approximately a 0.25-second delay on the camera video and robot commands, and about a 1-second delay on the 3D range data.

The controller is a Logitech WingMan RumblePad (see Figure 5), and of its features we use two analog thumb joysticks, a digital directional pad, six thumb buttons, and two finger "shoulder" buttons. This controller is very similar to present-day common video game controllers. The thumb joysticks control the motion of the robot arm, the directional pad controls the virtual view orientation, and the buttons control the virtual view zoom, operate the gripper, send the arm to stow position, and request a 3D snapshot.

Figure 5. Logitech WingMan RumblePad controller used to control the robot arm and user interface in the experiment.

4.2 Procedure

Before running the actual tests, each participant watched a tutorial video and performed a small training exercise to become familiar with the system. The tutorial included instructions to work quickly while trying to avoid collisions with the environment. The training exercises allowed the participant to explore how to change the view, how to control the robot arm in two different modes, and how to interpret the 3D range scan. After the tutorial, each participant collected at least one block with the whole system running the same as a normal experiment run, except the data was not recorded. Most participants spent about 15 minutes for the entire training portion. Participants were allowed to ask questions freely during the training. During the actual tests, however, only questions related to how to use the system were answered and not questions regarding the state of the world. Once finished with training, the users transitioned into the actual testing.

For the actual tests, each participant performed the task with 3 out of the 6 possible interfaces. The six variations are listed in Table 1. The participants progressed through the tests in pseudo-random, counterbalanced order to reduce order-dependent effects.

At the end of each test, a small survey had the users rate the interface on a scale of 1 to 10 with the following questions: 1. How much effort was required to use this interface effectively? 2. How difficult was it to learn this interface effectively? 3. How much confidence did you have in the robots actions? At the end of all tests, the users completed an additional survey with the following questions: 1. How much time in a week do you spend playing video games? 2. Rank the interfaces in order of your preference (selection from a list). 3. Any comments? At this point the experiment procedure was finished.

The independent variables in the experiment structure are (a) user interface display and (b) manipulator control mode. The user interface displays and manipulator control modes are explained in greater detail in Section 3.3.

Table 1. Variations in the user study. Parentheses show acronyms for each variant.

Interface/Control	*Joint control*	*End-effector control*
Video only	Variant 1 (VJ)	Variant 2 (VE)
3D scan only	Variant 3 (SJ)	Variant 4 (SE)
Video and 3D scan	Variant 5 (SVJ)	Variant 6 (SVE)

As shown in Table 1, variant 1 corresponds to the typical interface and control model. While not every traditional interface looks like this, we tried to capture the most common elements between these interfaces, namely the live video stream and current robot arm position. We left out some elements common to interfaces that are not relevant to the task, such as battery level indicators or robot "health status" information. The remaining variants of the interface use different arm control modes and the AV interface previously shown in Figure 3.

We measured operator performance primarily by time to complete tasks and number of collisions. To make these measurements, we recorded video footage from a separate camera behind the robot arm, then analyzed the video post-hoc to measure timings and collisions. See Table 2 for a listing of metrics and their measurement methods.

Table 2. Metrics and their respective methods for measurement in the user study.

Metric	Measurement
Manipulation time	Time from first arm movement to block deposit
Operator workload	Time performance, subjective survey
Situation awareness	Collisions
Quality of interactions	# interactions, time spent interacting
Operator preferences	Subjective survey

5 RESULTS

We present results for 30 people that participated in the user study. Twenty-four male and six female students at Brigham Young University participated. Participants were primarily young adults near age 18. Although we present results for 30 people, 34 participants actually attempted the experiment. One participant simply was not able to perform the task under any conditions, two participants were highly distracted, and one performed the task when a bug in the system had significantly slowed things down. Because we want as much as possible to avoid measuring such effects, we do not include data from these participants in the analysis. On three runs the system crashed after an hour of usage, though in each of these cases the participant was nearly finished, so we kept this data in the analysis. The primary results of interest that we found relate to (1) manipulation time, (2) collisions, and (3) subjective measures.

5.1 Manipulation Time

Participants performed the tasks fastest in the video-only conditions followed closely by the scan+video conditions, with scan-only slowest by a significant margin. People worked faster with joint control compared to end-effector control. See Figure 6 for the distributions of manipulation times, and Table 3 for significance between conditions. For the statistical significance analysis, we first use log correction to warp the data to a normal distribution. The D'Agostino-Pearson normality test shows that the warped data is significantly different from normal unless we remove statistical outliers (measured as 1.5 times the distance between quartiles 1 and 3), so we remove these outliers for significance analysis. The results in Table 3 show p-values for the corrected data.

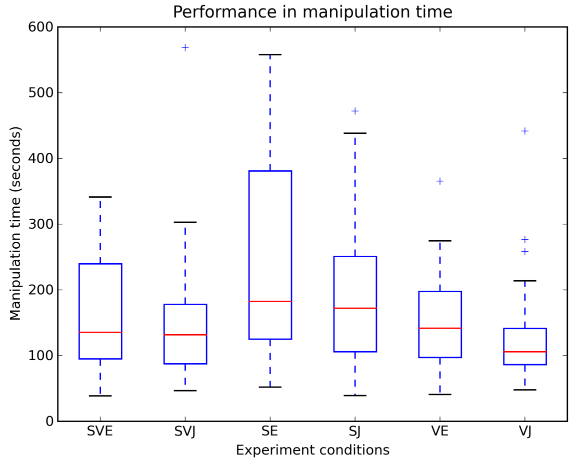


Figure 6. Performance measured as time required to perform a manipulation task. Note: To make this plot more readable, we do not show outliers above 600 seconds.

Table 3. Statistical significance analysis with p-values for two sample two-tailed t-tests. Bold numbers are significant at $\alpha = 0.05$. Note: S denotes 3D scan display, V denotes camera video display, E denotes end-effector control, and J denotes joint control.

Conditions	*SVE*	*SVJ*	*SE*	*SJ*	*VE*
VJ	**0.014**	0.080	**< 0.001**	**< 0.001**	**0.021**
VE	0.550	0.575	**< 0.001**	0.185	
SJ	0.574	0.065	**0.023**		
SE	**0.010**	**< 0.001**			
SVJ	0.287				

Several things might have caused the difference in time performance but we will discuss only a few. The likely largest factor in the slowness of the scan-only interfaces is misalignment between the 3D range scan and the virtual representation of the robot arm. In some cases, the misalignment was as much as 2 cm, which is substantial considering the clearance between the gripper opening and the block is about 1 cm on either side. Participants generally trusted the 3D range scan at first, but as they realized that the alignment was imperfect, they often grew frustrated in scan-only conditions. Because of the misalignment, users had to simply guess where and how the alignment deviates. On the other hand, with video conditions, users had undistorted visual feedback on the position of the gripper in relation to a target block, although depth perception is decreased.

Another factor in time performance involves view adjustments. When the 3D range scan was present, users spent some of the time adjusting the virtual perspective, while with the video-only interfaces almost no adjustments were made. Adjusting the view does not affect the viewpoint of the video camera, so people did not need to adjust the view much. One exception to this is in the end effector control mode, where the video is oriented corresponding to the rotation of the robot arm and the virtual viewpoint to give a sense for what direction the robot arm will move in the camera view. As such, users adjust the view somewhat in end effector control mode, although very little compared to when the 3D scan was present.

Another factor in time performance is the speed of the manipulator arm between control modes. In end-effector control the gripper moves at constant velocity in cartesian space regardless of position in the environment. On the other hand, joint control moves at constant angular velocity, which means that the farther the arm is extended, the faster the gripper moves in cartesian space. Participants tended to operate the arm in a more extended configuration, so end-effector control ended up moving somewhat slower overall.

5.2 Collisions

We separate collision types into two classes: (1) environment collisions and (2) target block collisions. Target block collisions are contact between the robot arm and the target blocks, except when a grasp action is underway. Environment collisions are contact between the robot arm and any part of the environment except target blocks. We report the total number of collisions for each condition across all participants for environment collisions in Figure 7 and for target block collisions in Figure 8. Because of high variability and a small sample size, this data has low statistical significance. We will look at statistical significance for collisions in future work.

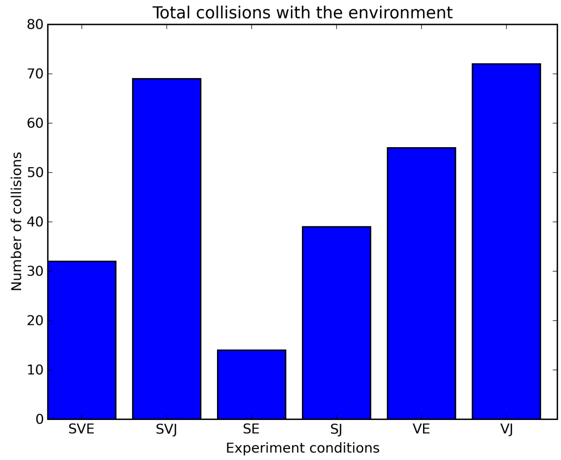


Figure 7. Total collisions with the environment.

In terms of displays, more environment collisions occurred with the video-only and scan+video interfaces than with the scan-only interfaces, while block collisions happened more with the scan-only interface than with any of the video interfaces. The alignment of the 3D scan is poor and coarse, so fine-grained, precise manipulation is difficult with scan-only interfaces. However, the granularity is suf-

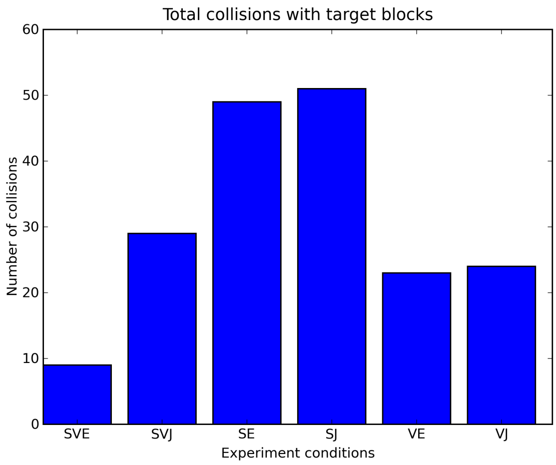


Figure 8. Total collisions with target blocks.

ficient for overall situation awareness and avoiding obstacles on a larger scale.

Even though the 3D scan is present in the scan+video interfaces, the video draws users' attention away from the 3D scan and so users appear to fail to notice imminent collisions [8].

Another factor in the number of collisions has to do with sensors. The webcam is closer to the environment than the SR-3000, so less of the environment is visible through the webcam. Such a view gives no information about what is to the sides of the camera, so users bump into things before they realize an obstacle is in the way. The webcam gives no depth information directly, so the user must use visual cues (for example shadows, relative sizes, and motion parallax) to determine depth.

When depositing a block with a 3D scan present, users often would drop the block well above the deposit box, while with video-only interfaces they usually had to make sure the block was inside the box before releasing the block. This reach-in approach generally caused several collisions with the deposit box, whereas the drop-from-above approach kept the arm clear of obstacles.

When we consider control type, more collisions occurred with joint control than with end-effector control. Joint control has a fixed angular resolution; that is, moving the control stick to the left rotates the robot arm a fixed number of degrees. On the other hand, end-effector control moves in cartesian space, so the resolution is finer as the arm extends as compared to joint control. Control resolution impacts the results because as the gripper nears an object, the user must finely adjust the position of the gripper for the final alignment.

We will present ANOVA for pairwise effects in future work.

5.3 Subjective Measures

In addition to objective measures for manipulation time and collisions, we also recorded a few subjective measures. Users answered the questions mentioned in Section 4.2. The results have sufficiently high variance that no conclusions can be made based on statistical significance, but we discuss trends anyway. For each of the subjective measures, we look at how many times a particular condition was rated better than the other conditions. Then we sum each of the condition votes for a global ranking. We look at subjective results for preference in overall interface, workload, learning required, and confidence in the robot's actions.

More users preferred joint control over end-effector control, except with the SVE condition. Users preferred variants with the 3D scan and video both present (SVE, SVJ). The preference ordering for all interfaces is as follows ($\sim$ denotes indifferent to and $\succ$ denotes preferred to): SVE $\sim$ SVJ $\sim$ VJ $\succ$ SJ $\succ$ SE $\succ$ VE. One possible reason for the low ranking of the scan-only interfaces (SE, SJ) is the misalignment between the virtual robot graphic and the 3D scan display. This disconnect led to frustration in the users, and that is reflected in this ranking. Video rotation (see Section 3.3) is probably the cause of dislike in the case of the video-only end-effector (VE) condition.

No obvious classification appears in the results for workload between the conditions. The preference ordering for workload is: SVE $\succ$ SJ $\succ$ VJ $\succ$ SVJ $\sim$ VE $\succ$ SE. Note that although performance was not the highest for the SVE condition, users apparently feel that the workload is lowest for the SVE condition. Strangely, SJ also ranks well for workload keeping in mind the misalignment between the 3D scan and robot arm graphic.

In terms of learning required to effectively operate the interface, joint control was generally easier to learn than end-effector control. The preference ordering for learning is: VJ $\succ$ SJ $\succ$ SVJ $\sim$ VE $\succ$ SVE $\succ$ SE. Joint control is most similar to many common types of control that people may have already been exposed to, such as driving a remote control car. End effector control is a new concept for most people, so more training is required to understand and use it effectively.

When it comes to understanding how the robot will move when given commands, users tended to feel more confident with joint control, except for the SVE condition. The preference ordering for confidence is: SVE $\succ$ VJ $\succ$ SJ $\succ$ SVJ $\succ$ VE $\succ$ SE. Part of the lack of confidence with end-effector control is likely due to the constraint for the target point. Normally it is best to constrain the target point to the robot arm's workspace, but due to implementation error we constrain it to a box that covers the robot's workspace. The side effect of such a constraint is that when the target point moves outside of the workspace, the robot's actions are confusing until the target point moves back inside the workspace. In the future it would be better to constrain the target point to the robot's workspace.

5.4 Discussion

The results seem to indicate that the 3D scan may actually increase workload, although it also improves situation awareness. Multiple perspectives with the 3D scan seem to provide better situation awareness, as fewer environment collisions happened when the 3D scan was present. In addition, it appeared from subjective observation that users who changed their view to check alignment had fewer collisions with the environment; future work should test this claim. Only a coarse understanding of the environment was possible with the resolution and calibration of the 3D scan, as indicated by a higher number of object collisions when video was absent. Without a good calibration for the robot arm and ranging camera (both intrinsic and exterior calibration), operators cannot trust the 3D scan to be accurate, and it may encourage them to depend more on camera video.

End-effector (grounded) control was poorly rated except when combined with the 3D scan and camera video even though fewer environment collisions happened with this control. This may be due in part to the view-dependent nature of the control. As an example case, we can consider when a user moved the gripper close to an object, and then changed the virtual viewpoint to gather more understanding. After the view change, the controls also change, and so when the user indicates the same direction on the controller as before, the robot moves differently. Apparently users think more from the robot's frame of reference than the computer screen's frame of reference. Perhaps more training with this control method would improve understanding, appeal, and performance.

Since the scan-only display with end-effector control exhibits the fewest world collisions and the scan+video display with end-effector control exhibits the fewest block collisions and has reasonable time performance, we predict that a better camera video integration with the 3D scan would reduce collisions while maintaining reasonable performance. This better integration could be done by virtually projecting the video in the 3D space near the virtual display of the robot arm, as if the virtual depiction of the camera was projecting the video.

6 CONCLUSIONS AND FUTURE WORK

Remote manipulation requires operators to control robots in complex environments with limited information. Typical user interfaces present information to the operator in the form of several camera video feeds and data displayed in several other channels. Although information required to perform tasks is available, the mental workload seems to be high enough that some of the information is forgotten or missed.

Evidence from the experiment suggests that although the ecological AV user interface for remote manipulation slows performance, it can also increase situation awareness and lessen mental workload. We believe that users were able to better understand the spatial relationship between the robot arm and its environment with the ecological interface compared to traditional interfaces. This represents a small step toward providing support for *mobile* manipulation. As this is an exploratory study, we have identified several problem areas with room for improvement. Future work needs to further test and refine these ideas.

In the future, we plan to make the display more integrated and ecological. We can add more autonomy to the robot arm, such as a "move close to that object" behavior. Because users spend much time adjusting the virtual perspective, we can improve the interaction method for adjusting the view with, for example, head tracking. Although we looked at operator workload with subjective measures and based on time performance, we can also use formalized methods such as NASA TLX or more objective measures such as secondary task performance.

ACKNOWLEDGEMENTS

We give thanks to Idaho National Laboratory and the employees there who supported this work. We also give thanks to ARL for support.

REFERENCES

[1] D. J. Bruemmer, D. A. Few, R. L. Boring, J. L. Marble, M. C. Walton, and C. W. Nielsen, 'Shared understanding for collaborative control', *IEEE Transactions on Systems, Man and Cybernetics, Part A*, **35**(4), 494–504, (July 2005).

[2] J. Casper and R.R. Murphy, 'Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center', *IEEE Transactions on Systems, Man and Cybernetics, Part B*, **33**(3), 367–385, (2003).

[3] Joseph Cooper and Michael A. Goodrich, 'Towards combining uav and sensor operator roles in uav-enabled visual search', in *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pp. 351–358, Amsterdam, The Netherlands, (2008). ACM.

[4] M. R. Endsley, 'Design and evaluation for situation awareness', in *In Proceedings of the Human Factors Society 32nd Annual Meeting*, pp. 97–101, Santa Monica, CA, (1988).

[5] M. A. Goodrich and Jr. Olsen, D. R., 'Seven principles of efficient human robot interaction', in *Proc. IEEE International Conference on Systems, Man and Cybernetics*, volume 4, pp. 3942–3948 vol.4, (2003).

[6] L. M. Hiatt and R. Simmons, 'Coordinate frames in robotic teleoperation', in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1712–1719, (9–15 Oct. 2006).

[7] D.B. Kaber, J. Riley, R. Zhou, and J.V. Draper, 'Effects of visual interface design, control interface type, and control latency on performance, telepresence, and workload in a teleoperation task.', in *In Proc. XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society*, (2000).

[8] R. Kubey and 79-86. M. Csikszentmihalyi, 286(2), 'Television addiction is no mere metaphor', *Scientific American*, **286**(2), 79–86, (2002).

[9] Jose A. Macedo, David B. Kaber, Mica R. Endsley, Preecha Powanusorn, and Seonwan Myung, 'The effect of automated compensation for incongruent axes on teleoperator performance', *Human Factors*, **40**(4), 541–553, (December 1998).

[10] Francois Michaud, Patrick Boissy, Helene Corriveau, Andrew Grant, Michel Lauria, Daniel Labonte, Richard Cloutier, Marc A. Roux, and David Iannuzzi, 'Telepresence robot for home care assistance', in *AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics*, (March 2007).

[11] P. Milgram and F. Kishino, 'A taxonomy of mixed reality visual displays', *IEICE Transactions on Information Systems*, **E77-D**(12), 1321–1329, (1994).

[12] P. Milgram, S. Zhai, D. Drascic, and J. Grodski, 'Applications of augmented reality for human-robot communication', in *Intelligent Robots and Systems '93, IROS '93. Proceedings of the 1993 IEEE/RSJ International Conference on*, volume 3, pp. 1467–1472 vol.3, (1993).

[13] L. A. Nguyen, M. Bualat, L. J. Edwards, L. Flueckiger, C. Neveu, K . Schwehr, M. D. Wagner, and E. Zbinden, 'Virtual reality interfaces for visualization and control of remote vehicles', *Autonomous Robots*, **11**(1), 59–68, (2001).

[14] Curtis W. Nielsen, *Using augmented virtuality to improve human-robot interactions*, Ph.D. dissertation, Brigham Young University, 2006.

[15] Bob Ricks, Curtis W. Nielsen, and Michael A. Goodrich, 'Ecological displays for robot interaction: A new perspective', in *Proceedings of IROS 2004*, Sendai, Japan, (2004).

[16] Paul S. Schenker, Eric T. Baumgartner, Sukhan Lee, Hrand Aghazarian, Michael S. Garrett, Randall A. Lindemann, D. K. Brown, Yoseph Bar-Cohen, Shyh-Shiuh Lih, Benjamin Joffe, Soon S. Kim, B. D. Hoffman, and Terrance L. Huntsberger, 'Dexterous robotic sampling for mars in-situ science', in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ed., David P. Casasent, volume 3208 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pp. 170–185. SPIE, (September 1997).

[17] Jean Scholtz, Mary Theofanos, and Brian Antonishek, 'Development of a test bed for evaluating human-robot performance for explosive ordnance disposal robots', in *HRI '06: Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pp. 10–17, New York, NY, USA, (2006). ACM Press.

[18] Thomas B. Sheridan, *Telerobotics, automation, and human supervisory control*, MIT Press, Cambridge, MA, USA, 1992.

[19] Katherine Tsui, Holly Yanco, David Kontak, and Linda Beliveau, 'Development and evaluation of a flexible interface for a wheelchair mounted robotic arm', in *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pp. 105–112, Amsterdam, The Netherlands, (March 2008). ACM.

[20] K.J. Vicente and J. Rasmussen, 'Ecological interface design: theoretical foundations', *Systems, Man and Cybernetics, IEEE Transactions on*, **22**(4), 589–606, (1992).

[21] C. D. Wickens and J. G. Hollands, *Engineering Psychology and Human Performance*, Prentice Hall, Upper Saddle River, NJ, 3rd edn., 2000.

[22] Christopher D. Wickens, John Lee, Yili D. Liu, and Sallie Gordon-Becker, *Introduction to Human Factors Engineering (2nd Edition)*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2003.

[23] D.D. Woods, J. Tittle, M. Feil, and A. Roesler, 'Envisioning human-robot coordination in future operations', *IEEE Transactions on Systems, Man and Cybernetics, Part A: Applications and Reviews*, **34**(6), 749–756, (November 2004).

Heuristic Rules for Human-Robot Interaction Based on Principles from Linguistics - Asking for Directions

Andrea Bauer and **Barbara Gonsior** and **Dirk Wollherr** and **Martin Buss** [1]

Abstract. Robots that are to assist humans in flexible and versatile ways, will not always possess all the information required to fulfill their task. Therefore robotic systems have to be able to retrieve information through natural language communication with humans. Natural language is often vague or ambiguous and thus hard to interpret by technical systems. To enable robotic systems to interpret natural language expressions correctly it is necessary to include findings from human-human communication into the dialog systems of robots. In this work the field of communication topics between human and robot is confined to the communication about space, more specifically to a robot asking a human for directions. This paper gives insight on theories from linguistics research focussing on asking for and giving directions. From these theories 10 heuristic rules for human-robot interaction are deduced, where 4 of the rules apply even to systems that are not able to communicate through natural language. Additionally a first experiment where a robot used the 4 basic heuristic rules to successfully ask passers-by for directions and find the way to an unknown goal location is presented.

1 Introduction

It is a goal of researchers around the world to design service robots that can be used by humans without technical training. This makes natural language communication robots a necessary ability for robotic systems. User-friendly robots should not only able to recognize speech and synthesize it, but also to understand the more complicated semantics. Therefore it is reasonable to include findings from linguistics in the form of heuristic rules for future dialog systems in robots.

Many examples of human-robot interaction (HRI) research exist where robots communicate with humans who have no experience with robotic systems, such as the museum guide robot Rhino [7] and its successor Minerva [23] which informed museum visitors about exhibitions. There are robots in shopping malls that relay useful precompiled information to humans, e.g. Toomas [11]. A communication robot for minor service tasks in hospitals, presented in [22], is intended to be used by people with little or no technical experience.

It can be assumed that systems which are to assist humans in a flexible and cognitive way will not always possess all the information required to fulfil their task. A central aspect of intelligent autonomous behavior is thus the ability to interact in order to retrieve information. As situations may often arise in which a robot will have to reach an unknown goal position our work focuses on HRI for asking for directions.

[1] Institute of Automatic Control Engineering (LSR), Technische Universität München, D-80290 Munich, Germany,
email: ab@tum.de, babsigonsior@gmx.de, dw@tum.de, mb@tum.de

Already there are robots communicating with humans about spatial information, such as [20, 12, 21]. A robot that retrieves spatial information from a grid map and translates it into linguistic spatial descriptions is presented in [20]. Human-augmented mapping [12] enables a robot to create a map of its environment by exploring it and asking a human to label the areas of interest for it, thus aligning a robotic map with human concepts. The robot Biron [21] integrates spoken dialog and visual localization to learn and label new places.

There are already some robots [18, 2, 17, 15] that ask humans for directions in simple settings and structured indoor environments. A wheelchair robot [18] can be given coarse qualitative route descriptions. The office robot Jijo-2[2] can learn the locations of offices and staff by moving around and asking humans for information. A robot asking for the way at a robotics conference is presented in [17]. A miniature robot that can find its way in a model town by asking for directions is described in [15].

As the ability to ask for directions seems to be a requirement in future robotic systems, we see the need in linguistic rules for dialog systems that will enable robots to ask humans for directions, reason with them about the given description and interpret the information correctly. Dialog systems in HRI are still mostly based on keywords, however as ”vagueness is one of the most salient, but also one of the most effective features of natural language” [13] there is more to understanding verbal information than to look up the lexical meaning of single words. Therefore it is necessary to include more complex rules in dialog systems. We propose 10 heuristic rules for HRI in an asking for directions situation.

The remainder of this paper is structured as follows. In Sec. 2 we give an overview of relevant theories from linguistics dealing with the problem of asking for directions. From these theories heuristic rules are derived and introduced in Sec. 3. An example of a robot following 4 of the proposed rules while interacting to retrieve direction information is presented in Sec. 4. Finally Sec. 5 provides concluding remarks.

2 Theoretical Background from Linguistics

Linguistics theories deal with the structures of complex verbal actions within human-human communication and the occurring psychological processes. Principles for reasoning about space can be found among them including the analysis of asking-for-directions dialogs and the different semantic meanings of verbal expressions or gestures. Theories relevant to the problem of asking for directions are the analysis of dialog structures and the complex deixis theory founded by Bühler [6] in 1934.

2.1 Dialog Structure

Wunderlich [24] analyzed asking-for-directions dialogs and found a common structure of four consecutive phases:

I *Introduction*: The asker addresses a respondent and defines the task, i.e. giving directions to a specified goal location, possibly defining the mode of transportation or other individual requirements.
II *Giving directions*: The respondent provides the necessary information by means of natural language and gestures, sometimes additionally with the help of a sketch.
III *Confirmation*: Either of the two partners confirms the information. In this phase further inquiries can be made.
IV *Conclusion*: The asker thanks the respondent and they part.

This schematic structure is very flexible, i.e. some phases may be interchanged or recur. Nevertheless it is a well-proven guideline for human-human interaction reflecting the intrinsic cognitive processes involved. One of these cognitive processes for the respondent is planning the description by building a cognitive map which is based on individual experiences. This cognitive map includes ”objects which are salient landmarks for nearly everybody” [14]. The respondent has to complete the task of separating these salient landmarks from individual experiences and present them to the asker. If this is achieved by using appropriate means of communication the asker will be able to build a corresponding cognitive map that represents the route to the goal location, structured by landmarks.

Spatial information in general and route information in specific is communicated using deictics, or deictic words, which are analyzed by deixis theory.

2.2 Deixis Theory

Deixis theory [6] is based on the assumption that communication acts can be assigned to two different fields of language, namely the symbolic field, comprising nouns which are symbols independent of the context, and the deictic field, that includes deictics, which vary depending on the context. In natural language communication referring is generally managed by deictics which point to subspaces of the deictic field and are the verbal equivalents to pointing gestures. In order to point to subspaces within the deictic field both speaker and listener have to share a common ”deictic space”. At the beginning of a face-to-face route description this deictic space is given through the range of the visual perception. In the course of the conversation the route description usually leaves the shared visual perception range and a new deictic space is built by the geographical knowledge of both partners. Another deictic space can be introduced by involving a map or a sketch that represents the real geographic space.

Referring to subspaces in a certain deictic space by using deictics depends on particular contextual factors such as the position of the speaker and the direction of gaze [14], can be summarized as the speaker’s origo.

2.3 The Origo

Bühler introduces the term ”origo” [6] which is conceptually conceived as the origin of a ”coordinate system of subjective orientation”. It is derived from the need of a ”basic reference point” [14] in a given deictic space, which includes three deictic dimensions [1]: the *personal dimension* including personal pronouns of the first and second person, the *spatial dimension* including demonstratives, adverbs, movement verbs, and prepositions, and the *temporal dimension* including temporal expressions. Thus the origo is defined by the personal mark ’I’, the spatial mark ’here’ and the temporal mark ’now’. Accordingly the origo is the reference point for personal, spatial and temporal deixis, which relates to other elements within these three deictic dimensions. Thus deictics can only be interpreted considering the origo.

Deictics have certain functions within an asking-for-directions dialog depending on their deictic dimension.

I *Personal deictics* define the actors within communication.
II *Spatial deictics* indicate the directions from landmark to landmark. They can be accompanied by gestures.
III *Temporal deictics* refer to the time domains of actions which have to be carried out to reach from one landmark to another. They are indicated by movement verbs.

The functions of deictics can be refined further on the basis of contextual factors and relations, as described by Fillmore [9] and Lyons [16]. As these refinements are beyond the scope of this paper, they are not further explained here.

In some cases deictics can only be located by using an additional ”relatum”, an object which is related to the origo [10]. For instance if it is referred to a subspace relative to a landmark, the landmark functions as a relatum relative to the perspective given by the origo. However if it is referred to a subspace relative to the speaker’s body, then the origo, or the speaker, functions as a relatum. If no new landmark is introduced after a deictic in a route description, the relatum is the origo of the previous landmark.

Problems and difficulties in interpreting deictics can occur, as deictics can be interpreted differently according to the deictic dimension, the context or whether or not they are used in combination with a relatum.

2.4 Problems in Interpreting Deictics

Several problems and difficulties that may occur while interpreting deictics and identifying the subspaces they refer to, have been identified by Klein [14].

I) Coordination problem: In a dialog situation all participants have their own origines. Since the origo is usually defined by the position and orientation of the current speaker, the listener must project that origo into her own system of orientation. As soon as the roles of speaker and listener are changed, the origo of the new speaker becomes essential and the other person has to adapt to it. Sometimes it is not clear whether the origo in a route description is the one of the speaker or the one of the listener. In the case of an asking-for-directions dialog the origines of the speaker and the listener are the same in terms of the position, as ’here’ encloses both speaker and listener. However, the orientations of the two communication partners differ and have to be coordinated, so as to avoid ambiguities of deictics.

II) Problem of the shifted origo: In the course of the asking-for-directions dialog both speaker and listener shift their origines into the perspective of an ”imaginary walker” [14] representing the addressee on her way projected into the future. A place that would be normally referred to as ’there’ may be called ’here’ within the route description, e.g. ’go straight until you see the park, here you need to turn left’. Thus the deictic ’here’ does not necessarily refer to the actual location but to the position of the shifted origo and must be interpreted depending on the context.

III) Problem with the use of an analogon: When humans use a sketch or a map to illustrate the described route they introduce a new deictic space. Consequently there are two deictic spaces involved, the map and the real geographic space represented by the map. The map functions as an analogon, where pointing to an element within it represents pointing to an element in the real space. Problems arise when the assignment of the two deictic spaces is not clear.

IV) Delimitation problem The subspace of the deictic space that a deictic points to can not be fully identified just by coordinating the origines of the dialog partners. The extend of a subspace is often vague and depends on context and environment, where the borders of the subspace must be established by gestures, verbal explanations or factual knowledge. For example, 'here' can be characterized as a subspace of the deictic space including the origo.

V) Problem of deictic oppositions: Deictic oppositions follow a certain system that, depending on the language can consist of two (e.g. in English: 'here' - 'there'), three (e.g. in German: 'hier' - 'da' - 'dort') or even more components. Examples of other deictic oppositions are the opposed meanings of 'left' and 'right' or 'this' and 'that'. Due to the fact that every communication partner has her own reference system there may emerge at least two different meanings for oppositional deictics. Their meaning depends on the context and additional information about the location in space and the delimitation or the interpretation of the shifted origo, where oppositional deictics may occur referring to subspaces in another deictic space, e.g. 'go straight on to the next intersection and turn right here'.

Even for humans it is sometimes hard to interpret deictics accurately, as any of these problems may occur in dialogs. For robots it is even harder to interpret deictics correctly, as they have to infer the right meaning from a series of single keywords retrieved by their speech recognition systems. In the following we propose guidelines in the form of heuristic rules that can be implemented in robotic systems to circumvent or solve the presented problems in HRI.

3 Heuristic Rules for Human-Robot Communication

Based on the theories from linguistics about human-human communication presented in the previous section we derive rules for the human-robot dialog of asking for directions. These rules should in hypothesis make the dialog more successful in terms of rendering it more natural and intuitive for the human partner, enabling the robot to interpret the possibly vague deictics correctly, and retrieve unambiguous route information that can be used by the robot to navigate.

Rules 1 to 4 are general rules for a robot asking for directions and can be applied by any interactive robot even if it is not equipped for natural language communication. Rules 5 to 10 are additional rules that apply to robots with natural-language-based dialog systems (text or speech). These rules provide guidelines that clarify and structure the vague deictics.

Heuristic Rule 1. *The dialog must be structured according to the four phases Introduction, Giving Directions, Confirmation, and Conclusion.*

This rule is a simple measure that adopts the structure from human-human communication to a human-robot dialog and therefore in hypothesis, should render the interaction more natural and intuitive for the human partners. Fig. 1 depicts the asking-for-directions dialog with the four phases, where the partner with the leading role is indicated in grey for each respective phase. It is the robot's task to

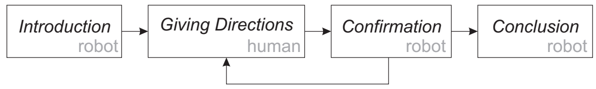


Figure 1. The structure of an asking-for-directions dialog. The partner with the leading role is indicated in grey in each phase.

initiate the *Introduction* phase, by addressing the human and introducing itself. A robot should introduce itself before asking for directions, as it is a strange novelty to humans in this situation. The *Introduction* phase is concluded by the robot when it asks for directions, where it has to specify how information can be given in accordance to the robot's abilities. The *Giving Directions* phase is initiated when the human begins to describe the route and it ends with the implicit or explicit declaration of the goal location. The *Confirmation* phase should be conducted by the robot which conveys to the human how the route description was interpreted. In this phase the robot should give feedback about the perceived information by displaying the perceived route in a bird's eye perspective including depictions of salient landmarks, solving the analogon problem presented in Section 2.4. This enables the human to see instantaneously that the robot has perceived the information correctly and if necessary to correct it. Lastly the *Conclusion* phase is initiated by the robot disclosing that no more information is needed. The robot should thank the human for the help and move on.

Heuristic Rule 2. *The robot must solve the coordination problem, by asking the human to specify the first direction with a pointing gesture.*

At the beginning of the *Giving Directions* phase it is not immediately clear whether directions are given relative to the origo, i.e. the personal reference system, of the human or the robot. As outlined in Fig. 2, a point of interest (PoI) is described as being located in opposite directions relative to the reference system of the human O_h or the robot O_r, rendering the direction information ambiguous.

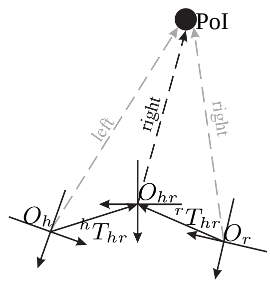


Figure 2. A PoI is referred to differently from the personal reference systems, or origines, of two dialog partners O_h and O_r. After coordinate transformations, ${}^rT_{hr}$ and ${}^hT_{hr}$ respectively, both partners use a common reference system O_{hr} to refer to a PoI in an unambiguous way.

To solve this problem the human is asked to point in the first direction of the route, as the gesture is unambiguous. At that point further directions are given relative to this direction. In mathematical terms, this pointing introduces a common reference system, O_{hr}, which is oriented towards the indicated direction, and specifies the necessary transformations, ${}^rT_{hr}$ and ${}^hT_{hr}$ of the reference systems O_h and O_r.

From then on it is understood that further directions are given relative to the same orientation. This rule solves the coordination problem, and is crucial for the success of the whole dialog.

Heuristic Rule 3. *The robot must break the perceived route description down into route segments and save the information as a route graph.*

The robot needs an internal representation of the route information to be able to reason about it and to navigate by it. Therefore the perceived route description must be broken down into route segments and stored as a route graph $G\langle\text{nodes, edges}\rangle$. Nodes $N_i(t_j)$ of the route graph represent landmarks of type t_j in the route description. Edges $E_i(a_i, N_k, N_l)$ represent actions a_i connecting the respective nodes, i.e. landmarks, N_k and N_l. An implementation of a system building a topological route graph from human direction information is presented in [3].

Heuristic Rule 4. *The robot must interpret the basic directions 'left', 'right', 'straight', and 'back'.*

The direction deictics 'left', 'right', 'straight', and 'back' establish the directions of actions starting at the last landmark. These deictics are the most important and basic ones in an asking-for-directions dialog. Thus the minimal requirement for a dialog system to interpret route directions is to be able to interpret these four deictics.

A possible definition for the deictics 'left', 'right', 'straight', and 'back' is to interpret every deictic as an angular range of 90°, as depicted in Fig. 3. An example of this definition is given in [8]. The definition could be refined for example by including expressions like 'diagonally left'.

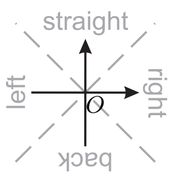


Figure 3. The deictics 'left', 'right', 'straight', and 'back' as angular ranges of 90°.

Heuristic Rule 5. *The robot must identify landmarks in the route description.*

Landmarks structure the route description, provide starting and end points of route segments and they must be identified when they occur in the description. A node $N_i(t_j)$ must be inserted into the route graph for every landmark in the route description. There are four different ways to identify landmarks in a route description. Firstly, they can be identified through previous knowledge saved in the dictionary of the dialog system. Secondly, when the human points to a landmark it is recognized from gestural and visual clues. Thirdly, it can be identified from a non-deictic description of the landmark, e.g. 'the yellow building'. Fourthly, a landmark can be identified by its relationship to previous landmarks, as in 'the third intersection', where the next one or more landmarks are of the same type as the explicitly mentioned landmark. In this case the robot must insert the corresponding number of nodes and edges into the route graph.

The last landmark in the route description is the goal location, or if the route is very complex, it may be a preliminary goal.

Heuristic Rule 6. *The robot must interpret 'here' and 'there' depending on when they occur in the route description.*

If the deictics 'here' or 'there' accompany a pointing gesture at the beginning of a route description, these words define the first landmark, namely the spatial structure that the gesture points to.

If these deictics occur later on in the route description they are usually linked with landmarks, i.e. they either stand in the place of landmarks or accompany them. Thus they structure the route description just as landmarks do. When one of these deictics occurs a new route segment begins in the description, therefore a new edge $E_i(a_i, N_k, N_l)$ must be inserted into the route graph.

Heuristic Rule 7. *The robot must identify movement verbs.*

Movement verbs define the actions that connect the landmarks and must be identified by the robot. For every movement verb an edge $E_i(a_i, N_k, N_l)$ must be inserted into the route graph. All movement verbs can all be interpreted as the one neutral movement verb 'move'. In a route description a movement verb may even be omitted as it may be clear from the context.

Heuristic Rule 8. *Deictics must be classified according to their semantic attributes.*

The deictics input must be classified by the robot according to semantic attributes in order to simplify the interpretation of the expression. To this end the semantic attributes of deictics must be stored in the dictionary of the dialog system. Relevant semantic attributes of deictics are the *deictic dimension*, the *distance range*, the *delimitation*, and whether or not a *relatum* is needed. The deictic *dimension* is either *spatial*, *temporal*, or *personal* and denotes whether a deictic points to a region in space, in time, or to a person. The *distance* of a deictic denotes how far the pointed region is away from the origo. It can be sub-classified into *origo-inclusive* and *origo-exclusive*, where the latter can be *close* to the origo or *far* away from it. The *delimitation* expresses whether or not the region a deictic points to has clearly defined boundaries. The last attribute indicates whether a deictic needs a *relatum* or whether it can occur without one. Deictics which can not be classified into origo-inclusive or origo-exclusive express only an uncertain proximity or distance, and therefore require a relatum to define the location in the deictic space [10]. As a general rule all personal deictics and deictics which are delimited do not necessarily require a relatum. Table 1 gives an overview of the most common deictics with the relevant semantic attributes, it does not provide an exhaustive list of deictics, but gives the reader an idea of how to classify deictics.

As a consequence of saving the attributes the interpretation of deictics is facilitated. The deictic dimension reveals whether or not a deictic is part of a route description, in the case of spatial and temporal deictics, or not, in the case of personal deictics. The distance provides an approximation of the spatial location of the described feature, as does the delimitation. Finally a deictic that needs a relatum must be interpreted together with the relatum.

Heuristic Rule 9. *To simplify the interpretation of deictics, the temporal domain can be mapped to the spatial domain.*

In a route description the passing of time corresponds to traveling along the route. When giving directions the 3-dimensional space is projected to a simple (1-D) path, where landmarks confine route segments according to heuristic rule 3. The temporal domain is also 1-dimensional where landmarks confine certain periods of traveling the path. Thus time and space have a similar structure in a route description and can be mapped to one another and the number of expressions that have to be analyzed further is reduced. Temporal expressions in

Table 1. Common deictics with relevant semantic attributes.

DEICTIC	DIMENSION			DISTANCE			DELIMITATION		NEEDING RELATUM	
	spatial	temporal	personal	origo-inclusive	close	far	yes	no	yes	no
'here'	×			×				×		×
'there'	×				×	×		×		×
'near'	×				×			×	×	
'far from'	×					×		×	×	
'left'	×							×	×	
'right'	×							×	×	
'in front of'	×							×	×	
'behind'	×							×	×	
'this'	×				×		×			×
'that'	×					×	×			×
'now'		×		×				×		×
'then'		×			×	×		×	×	
'soon'		×			×			×		×
'later'		×				×		×		×
'I'			×	×			×			×
'you'			×		×	×	×	×		×

route descriptions occur only in present and future tense, the most common of these are listed below.

'now'	$\rightarrow$	'here'
'then'	$\rightarrow$	'there'
'soon'	$\rightarrow$	'near'
'later'	$\rightarrow$	'far'
t	$\rightarrow$	$x = \frac{v}{t}$, with walking velocity v

If a temporal deictic cannot be mapped to a spatial equivalent, it can still be interpreted as a start or end point of an action.

Note that 'then' following 'if' is not temporal but causal and must not be mapped to the spatial domain.

Heuristic Rule 10. *The delimitation problem must be solved by modeling the distance ranges of various expressions such as 'here', 'near', and 'far' depending on the environment.*

It is clear that the deictic 'here' denotes the position of the actual or shifted origo however it is unclear how far the region 'here' stretches. The deictics 'near' and 'far' denote regions in space that do not include the origo and therefore do not coincide with the region 'here', where it is clear that the first expression denotes a region closer to 'here' and the latter expression one that is farther away. The reaches of the spatial deictics, such as 'here', 'near', and 'far', depend on subjective perception, the personal viewpoint and on the context, such as the used mode of transportation and the extension of the described route. The probability density functions (pdf's) of the deictics 'here' $p_{\text{'here'}}(d)$, 'near' $p_{\text{'near'}}(d)$, and 'far' $p_{\text{'far'}}(d)$ are estimated based on the results of a user study, presented in [4]. Fig. 4 depicts those pdf's depending on the distance d in an urban environment. The pdf's are modeled as

$$p_{\text{'here'}}(d) = \frac{1}{\mu} e^{\frac{d}{\mu}} \text{ , with } \mu = 148.7 \text{ ,}$$

$$p_{\text{'near'}}(d) = \frac{1}{d\sigma\sqrt{2\pi}} e^{\frac{-(ln(d)-\mu)^2}{2\sigma^2}} \text{ , with } \mu = 5.8 \text{ , } \sigma = 0.51 \text{ ,}$$

$$p_{\text{'far'}}(d) = \alpha\, \beta\, d^{\beta-1}\, e^{-\alpha\, d^b} \text{ , with } \alpha = 952.1 \text{ , } \beta = 5.85 \text{ ,}$$

where $p_{\text{'here'}}(d)$ is an exponential function, $p_{\text{'near'}}(d)$ is a lognormal function, and $p_{\text{'far'}}(d)$ is a Weibull function.

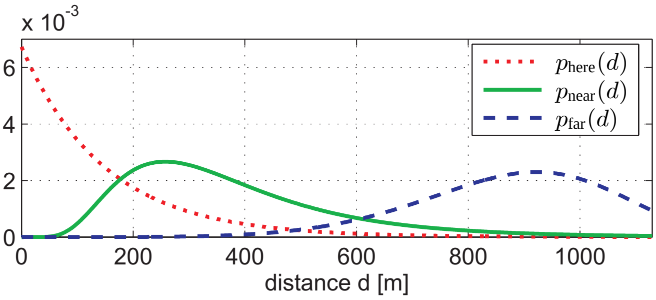


Figure 4. Estimated pdf's for the terms 'here', 'near', and 'far' over the distance d [m] from the origo in an urban environment.

The deduced heuristic rules present a guideline to researchers implementing the ability of asking for directions in a robotic system. Depending on the hardware and other constraints they may be altered or expanded to meet the requirements. The first 4 basic rules have been implemented on a robot and tested during a HRI experiment.

4 Experiment

In a first experiment previously described in [5] 4 of the proposed rules (heuristic rules 1 to 4) were included in the dialog system of a robot that had to ask passers-by for directions in order to reach an unknown goal location in an urban environment, without the use of GPS or previous map knowledge. Only the first 4 general heuristic rules were included as the robot was not equipped for natural language communication.

4.1 The Autonomous City Explorer

The *Autonomous City Explorer (ACE)* robot as depicted in Figure 5 was equipped with an active-stereo-camera head for human tracking and gesture recognition, an animated mouth, a loud speaker, and a touch screen for HRI, as well as a differential wheel mobile platform and laser range finders for navigation.

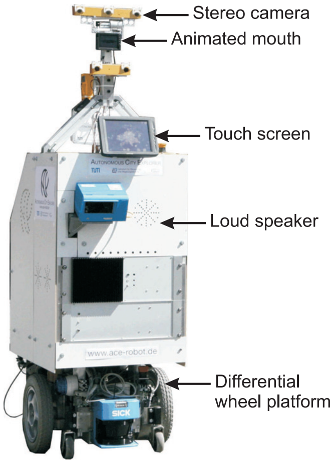


Figure 5. The robot *ACE* with principal hardware components.

The robot communicated with the human firstly by speaking over a loud speaker (English or German language) using MaryTTS [19] for speech synchronization. A mouth displayed on a small monitor was animated synchronously to the speech. Additionally the robot could present images and text to the human on a touch screen. As robustness to environmental disturbances such as noise was an important requirement for the system, the touch screen, was also used as the main means of input from the human. The robot was able to recognize and interpret pointing gestures, as described in [5].

4.2 Heuristic Rules Included in the HRI

As the completion of the task, i.e. to navigate to an unknown goal location in an urban environment, depends on a successful interaction, the design of the dialog was a crucial component and the first 4 of the proposed heuristic rules were included. The included rules are all general guidelines for asking for directions and not specific for natural language communication, therefore they could be applied to the robot that does not understand natural language but presents the human with options for answers.

Firstly when the robot addresses a human it introduces itself according to heuristic rule 1 and the dialog is structured in four phases, where the *Giving Directions* phase coincides with the *Confirmation* phase. The robot depicts every direction information on the touch screen whenever it is given. This allows the human to check whether the robot has interpreted the input accurately and if necessary to correct it. At the beginning of the *Giving Directions* phase the robot asks the human to point in the direction it has to go into first. So that at this point human and robot align their personal reference systems, i.e. origines, and it is understood that further directions are given relative to the common reference system. The human is then ask to give further directions through buttons on the touch screen, where buttons for the basic directions are provided, according to rule 4.

An exemplary dialog is given in the following:

ACE: "Hello, my Name is ACE. I need to get to Marienplatz."
Human confirms by touching screen
ACE: "Please look into my eyes and point in the direction I should take now, as shown on the screen."
Human points
ACE turns its head, takes an image of the indicated direction, and displays it on the touch screen
ACE: "Please indicate the exact position in the image."
Human touches the screen in the corresponding place
ACE: "Please tell me how I should proceed then."
Human gives direction by pressing buttons representing directions and indicates when the route instruction is complete
ACE: "Thank you very much for your help."
ACE: "Please step back, I will go now."

Following heuristic rule 3 the robot builds an internal representation of the route as a topological route graph with nodes representing intersections and edges representing actions connecting the intersections.

4.3 Experimental Results

The *ACE* robot found its way to a designated goal location without the help of GPS data or previous map knowledge, solely by asking passers-by for directions. The robot managed to travel the distance of $1.5\,km$ between the campus of Technische Universität München and Marienplatz (the central square of Munich) in 5 hours, interacting with 38 passers-by. The average duration of an interaction was 1.5 minutes. An example of an interaction between *ACE* and a passer-by is shown in Figure 6, where a passer-by points into the first direction the robot has to go while ACE follows the gesture with the camera head to take an image and present it to the human.

Figure 6. *ACE* following the pointing gesture of a passer-by with the camera head.

Figure 7 depicts the route graph, built by the robot, along with a corresponding occupancy grid in a satellite image of the real environment. The robot was currently positioned at node N_2, where it had retrieved topological knowledge of the route that lied ahead (white). The direction from N_2 to N_3 was given through a gesture, further directions were given through buttons on a touch screen.

The fact that the robot reached its goal solely with the help of instructions of passers-by who were not previously instructed on how to interact, leads to the conclusion that the interactions were successful and therefore intuitive for the humans. Problems arose, where the human partners had too high expectations of the abilities of the robot. For example many users expected the robot to be able to understand speech at first and tried to answer through natural language until they realized that they had to use the touch screen to communicate. Also some humans had problems making a pointing gesture that was recognizable by the robot, as the robot could only recognize gestures where the arm was fully extended and therefore instructed the humans on how to point. The robot was sent in the wrong direction

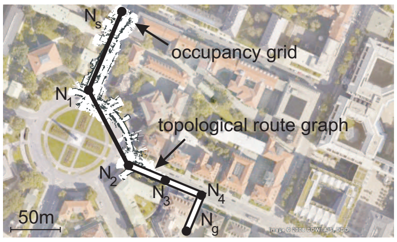


Figure 7. Example of a route graph retrieved from human instructions.

once by a passer-by, but this wrong information was corrected by the next interaction partner after the robot had to stop because the path was blocked. Otherwise no conflicting information occurred.

To summarize most of the problems in HRI arised when it was not entirely clear to the human, what abilities the robot possessed and what the constraints were. Therefore the robot must clarify those points at the beginning of every interaction. The next step is to make the communication more natural, in specific improve the gesture recognition and include a robust speech recognition system incorporating all of the introduced heuristic rules for interpreting deictics.

5 CONCLUSION

In future applications we see the necessity for robots to be able to retrieve the missing information through HRI. Human-human communication was used as a role model for human-robot communication. To this end theories from linguistics research focussing on asking for and giving directions were presented. Heuristic rules for a robot asking a human for the way were deduced. Additionally we presented an experiment where a robot used 4 of the proposed rules to successfully asks passers-by for directions and find an unknown goal location in Munich, Germany.

The heuristic rules that have been implemented so far proved useful in rendering the interaction more intuitive for the human, helped avoiding ambiguities in the route descriptions, and enabled the robot to build an internal representation of the route description. The experiment showed that it is necessary to explain exactly to the human partners how to interact with the robot and what the constraints are. Our future work will be to include all of the proposed heuristic rules in a text based dialog system in order to evaluate them.

ACKNOWLEDGEMENTS

This work is supported in part within the DFG excellence initiative research cluster *Cognition for Technical Systems – CoTeSys*, see also `www.cotesys.org` and within the European Commission as part of the *Robot@CWE project*, see also `www.robot-at-cwe.eu`.

REFERENCES

[1] S. Anderson and E. Keenan, *Language typology and syntactic description, Vol. 3: Grammatical categories and the lexicon*, chapter Deixis, 259–308, Cambridge University Press, 1985.

[2] H. Asoh, Y. Motomura, F. Asano, I. Hara, S. Hayamizu, N. Vlassis, and B. Kröse, 'Jijo-2: An office robot that communicates and learns', *Intelligent Systems*, **19**(5), 46–55, (2001).

[3] A. Bauer and M. Buss, 'Information retrieval system for human-robot communication - asking for directions', in *ICRA*, (2009).

[4] A. Bauer and M. Buss, 'Modeling probability distributions of deictic expressions 'here', 'near', and 'far'', Technical report, Inst. of Automatic Control Eng., Technische Universität München, (2009).

[5] A. Bauer, K. Klasing, G. Lidoris, Q. Mühlbauer, F. Rohrmüller, S. Sosnowski, T. Xu, K. Kühnlenz, D. Wollherr, and M. Buss, 'The autonomous city explorer: Towards natural human-robot interaction in urban environments', *International Journal of Social Robotics*, (2009).

[6] K. Bühler, *Theory of language. The representational function of language*, Benjamins, 1934/1990.

[7] W. Burgard, A. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, 'Experiences with an interactive museum tour-guide robot', *Artificial Intelligence*, **114**(1-2), 3–55, (1999).

[8] B. Claus, K. Eyferth, C. Gips, R. Hornig, U. Schmid, and F. Wysotzki, 'Reference frames for spatial inference in text understanding', in *Spatial Cognition*, pp. 241–266. Springer, (1998).

[9] C. Fillmore, *Lectures on Deixis*, U. of Chicago Press, 1997.

[10] E. Fricke, 'Origo, pointing, and speech: the impact of co-speech gestures on linguistic deixis theory', *Gesture*, **2**(2), 207–226, (2002).

[11] H. Gross, H. Böhme, C. Schröter, S. Müller, A. König, C. Martin, M. Merten, and A. Bley, 'Shopbot: Progress in developing an interactive mobile shopping assistant for everyday use', in *SMC*, (2008).

[12] H.Hüttenrauch, K.Severinson Eklundh, and E.A.Topp A.Green, 'Investigating spatial relationships in human-robot interaction', in *International Conference on Intelligent Robots and Systems*, (2006).

[13] W. Klein, *Spatial Orientation: Theory, Research, and Application*, chapter Deixis and spatial orientation in route directions, 283–320, New York: Plenum, 1980.

[14] W. Klein, *Speech, Place, and Action: Studies in Deixis and Related Topics*, chapter Local Deixis in Route Directions, 161–182, New York: Wiley, 1982.

[15] S. Lauria, G. Bugmann, T. Kyriacou, and E. Klein, 'Instruction based learning: how to instruct a personal robot to find hal.', in *European Workshop on Learning Robots*, (2001).

[16] J. Lyons, *Linguistic Semantics: An Introduction*, Cambridge University Press, 1995.

[17] M. Michalowski, S. Sabanovic, C. DiSalvo, D. Busquets Font, L. Hiatt, N. Melchior, and R. Simmons, 'Socially distributed perception: Grace plays social tag at aaai 2005', *Autonomous Robots*, **22**(4), 385–397, (2007).

[18] R. Müller, T. Röfer, A. Lankenau, R. Musto, K. Stein, and A. Eisenkolb, 'Coarse qualitative descriptions in robot navigation', in *Spatial Cognition II*, pp. 265–276, (2000).

[19] M. Schröder and J. Trouvain, 'The german text-to-speech synthesis system MARY: A tool for research, development and teaching', *International Journal of Speech Technology*, **6**, 365–377, (2003).

[20] M. Skubic, D. Perzanowski, S. Blisard, A. Schultz, W. Adams, M. Bugajska, and D.Brock, 'Spatial language for human-robot dialogs', in *Transactions on Systems, Man and Cybernetics*, pp. 154–167, (2004).

[21] T. Spexard, S. Li, B. Wrede, J. Fritsch, G. Sagerer, O. Booij, Z. Zivkovic, B. Terwijn, and B. Kröse, 'Biron, where are you? enabling a robot to learn new places in a real home environment by integrating spoken dialog and visual localization', in *IROS*, pp. 943–940, (2006).

[22] D. Spiliotopoulos, I. Androutsopoulos, and C. Spyropoulos, 'Human-robot interaction based on spoken natural language dialogue', in *E. Workshop on Service and Humanoid Robots*, (2001).

[23] S. Thrun, M. Beetz, M. Bennewitz, W. Burgard, A.B. Cremers, F. Dellaert, D. Fox, D. Haehnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, 'Probabilistic algorithms and the interactive museum tour-guide robot minerva', *IJRR*, **19 (11)**, 972–999, (2000).

[24] D. Wunderlich, *Linguistische Berichte*, volume 53, chapter Wie analysiert man Gespräche? Beispiel: Wegauskünfte, 41–76, Helmut Buske Verlag, 1978.

A dynamic field approach to goal inference and error monitoring for human-robot interaction[1]

Estela Bicho[2] and **Luís Louro**[3] and **Nzoji Hipólito**[4] and **Wolfram Erlhagen**[5]

Abstract. In this paper we present results of our ongoing research on non-verbal human-robot interaction that is heavily inspired by recent experimental findings about the neuro-cognitive mechanisms supporting joint action in humans. The robot control architecture implements the joint coordination of actions and goals as a dynamic process that integrates contextual cues, shared task knowledge and the predicted outcome of the user's motor behavior. The architecture is formalized by a coupled system of dynamic neural fields representing a distributed network of local but connected neural populations with specific functionalities. We validate the approach in a task in which a robot and a human user jointly construct a toy 'vehicle'. We show that the context-dependent mapping from action observation onto appropriate complementary actions allows the robot to cope with dynamically changing joint action situations. This includes a basic form of error monitoring and compensation.

1 INTRODUCTION

As robot systems are moving as assistants into human everyday life, the question how to design robots capable of acting as sociable partners in collaborative joint activity becomes increasingly important ([4], [12]). Useful and efficient human-robot interaction requires that both teammates coordinate and synchronize their actions and decisions in a shared task. In order to decrease the workload of the human and to increase user satisfaction, the robot should equally contribute to this coordination effort. This necessarily means that the robot should be endowed with cognitive capacities such as action understanding and goal inference. Humans achieve their remarkable fluent organization of joint action by anticipating the intentions of others [21]. In our everyday social interactions we continuously monitor the actions of our partners, interpret them effortlessly in terms of their outcomes and use these predictions to select adequate complementary behaviours. Very often this happens without the need for explicit verbal communication. Imagine for the instance the joint action task of preparing a dinner table. The way how a partner grasps a certain object, e.g., a coffee cup, transmits to the observer important information about the ultimate goal of the action. Depending on the grip type, the partner may want to place the cup on the table or, alternatively, has the intention to hand it over. Being able to predict the goal of the whole action sequence at the time of the grasping allows the observer to timely prepare for receiving the cup, or to initiate the selection of another object for the dinner table.

This paper presents our ongoing research towards creating socially intelligent robots that are able to flexibly adjust their goal-directed behaviours in dependence of the predicted outcomes of actions of their human partners [3]. Our approach is heavily inspire by recent experimental and theoretical findings about the neuro-cognitive mechanisms underlying joint action in humans and other primates ([18], [25]). We believe that designing cognitive control architectures on the basis of these mechanisms defines a very promising research direction to reduce the significant imbalance in social and cognitive skills between human and robot that still exists today. Ultimately, implementing a human-like joint action model in the robot will contribute to more natural HRI since the teammates will become more predictable for each other. This in turn will increase the acceptance by humans. A recent HRI user study with a simulated robotic teammate revealed that anticipatory action selection seems to be a natural expectation of a robotic assistant in known joint action tasks [17]. The robot is perceived as a full partner that contributes to the team's fluency and success only if it acts in anticipation of the needs of the human user.

Several neuro-cognitive mechanisms that are believed to underlie successful human joint action define fundamental components of the robot control architecture. An impressive body of experimental evidence from studies investigating action and perception in a social context suggests that motor simulation routines in the brain support the understanding of other's actions and facilitate overt imitation [25]. The fundamental idea is that perceived actions are automatically mapped onto corresponding motor representation of the observer to predict or replicate the action effect. Over the last couple of years, the suggested close perception-action link has inspired robotics work mainly in the domain of learning by imitation and social development (e.g., [5], [1], [10], [15], [19]). For implementing a high-level goal inference capacity in the context of HRI, it is important that the matching takes place on a sufficiently abstract level related to the goal or desired end state of an action sequence. This allows the robot to predict actions of the teammate despite the obvious differences in embodiment and motor skills between human and robot. However, for action selection in cooperative tasks an automatic and direct resonance of matching motor structures is in general not beneficial. Normally, action observation should facilitate the selection of a non-imitative, complementary behaviour. Moreover, to cope with dynamically changing joint action conditions, the decision about what defines the most adequate complementary action

[1] The present research was conducted in the context of the fp6-IST2 EU-project JAST (proj.nr. 003747) and partly financed by the FCT grants POCI/V.5/A0119/2005 and CONC-REEQ/17/2001.
[2] Dept of Industrial Electronics, University of Minho, Portugal. Email: estela.bicho@dei.uminho.pt.
[3] Dept of Industrial Electronics, University of Minho, Portugal. Email: llouro@dei.uminho.pt.
[4] Dept of Industrial Electronics, University of Minho, Portugal. Email: nhipolito@dei.uminho.pt.
[5] Dept of Mathematics for Sciences and Technology, University of Minho, Portugal. Email: wolfram.erlhagen@mct.uminho.pt.

should depend on additional contextual cues. Recent evidence from neurophysiological and behavioural studies shows that the automatic mapping from action observation onto action execution is indeed more complex and flexible as previously thought ([18], [23]). The robot control architecture reflects these findings by implementing a context-dependent mapping that is biased by the inferred goal of the human user.

As a theoretical framework for the high-level control of the robot we have used the Dynamic Neural Field (DNF) approach to robotics [8]. Originally introduced as a simplified mathematical model for pattern formation in neural populations [2], DNFs have been later generalized and applied to the cognitive domain (for a recent review see [20]). The architecture of DNFs reflects the hypothesis that strong recurrent interactions in local populations of neurons form a basic mechanism for cortical information processing. These interactions support the existence of self-stabilized inner states that allow the cognitive agent for instance to compensate for temporally missing sensory input, or to anticipate future environmental inputs that may inform the decision about a specific goal-directed behaviour. The DNF-based model for joint action consists of a distributed network of reciprocally connected neural populations that represent in their activation patterns specific task-relevant information. It implements the idea that the coordination of actions and decisions among the teammates is a dynamic process that builds on the continuous integration of input from representations of the inferred goal of observed actions, contextual cues and shared task knowledge. The representation of the complementary action that gets the strongest support from all connected populations will win the dynamic competition process among all possible actions.

The dynamic field architecture has been validated in a joint construction task in which the human-robot team assembles a toy 'vehicle' from its components knowing the construction plan. The study differs from conceptually related HRI work [17] in the sense that the robot is not only serving the user (e.g., holding out pieces for the user) but is able to perform itself the assembly task. This symmetric situation challenges the joint coordination of decisions and actions. The focus of the results reported here is on successful trials in which the robot shows anticipatory action selection. Since coordination and other errors may occur even in tasks that are well known to the teammates [22], performance monitoring and error detection is another topic that we have addressed. In the present implementation, complementary action selection in error trials may range from simple head nodding to pointing. In addition, we have integrated and tested a speech production system that allows the robot not only to explain the error in some more detail but to send in general feedback about its reasoning to the user.

The paper is organized as follows: Section 2 introduces the joint construction task and the robotic platform. Section 3 gives an overview about the cognitive control architecture. Section 4 presents the basic concepts of the dynamic field framework. The results of the human-robot interactions are described in section 5. The paper ends with a discussion of concepts, results and a short outlook.

2 JOINT CONSTRUCTION TASK

To validate the dynamic field architecture for human-robot interaction we have chosen the joint construction of a toy 'vehicle'from components that are initially distributed on a table. (Figure 1). The task requires only a limited number of different motor actions to be performed by the team but is complex enough to show the impact of action monitoring and evaluation on action selection. The components that have to be manipulated by the robot were designed to limit the workload for the vision and the motor system of the robot. The toy object consists of a round platform with an axle on which two wheels have to be attached and fixed with a bolt. Subsequently, 4 columns have to be plugged into holes in the platform. The placing of another round object on top of the columns finishes the task. It is assumed that each teammate is responsible to assemble one side of the toy. Since the working areas of the human and the robot do not overlap, the spatial distribution of components on the table obliges the team to coordinate in addition handing-over sequences. It is further assumed that both partners know the construction plan and keep track of the subtasks which have been already completed by the team. Since the desired end state does not uniquely define the logical order of the construction, at each stage of the construction the execution of several subtasks may be simultaneously possible. The main challenge for the team is thus to efficiently coordinate in space and time the decision about actions to be performed by each of the teammates.

Figure 1. Human-robot team for the joint construction of a toy 'vehicle'. The vehicle consists of a (red) round platform with an axle where two (green) wheels have to be mounted and fixed with (magenta) bolts.

For the HRI experiments we used a robot built in our lab. It consists of a stationary torus on which a 7 DOFs AMTEC arm (Schunk GmbH) with a 3-fingered BARRET hand (Barrett Technology Inc.) and a stereo camera head are mounted. A speech synthesizer (Microsof Speech SDK 5.1) allows the robot to communicate the result of its reasoning to the human user. For the control of the arm-hand system we applied a global planning method in posture space that allows us to integrate optimization principles derived from experiments with humans [10]. The information about object type, position and pose is provided by the camera system. The object recognition combines color-based segmentation with template matching derived from earlier learning examples [24]. The same technique is also used for the classification of object-directed, static hand postures such as grasping and communicative gestures such as pointing or demanding an object.

3 COGNITIVE CONTROL ARCHITECTURE

Figure 2 presents a sketch of the multi-layered robot control architecture for dynamic decision making and performance monitoring in joint action that is based on known neuro-cognitive mechanisms. Ultimately, the architecture implements a context-dependent mapping between observed action and executed action. The fundamental idea is that the mapping takes place on the level of abstract motor primitives defined as whole goal-directed motor acts

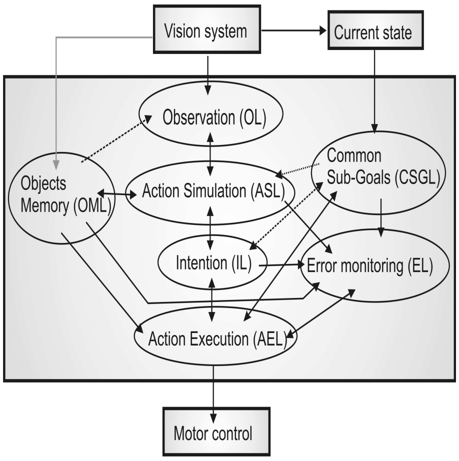


Figure 2. Cognitive control architecture for joint action. It implements a mapping from observed actions (layer OL) onto complementary actions (layer AEL) taking into account the inferred action goal of the partner (layer IL), detected errors (layer EL), contextual cues (layer OML) and shared task knowledge (layer CSGL). The goal inference capacity is based on motor simulation (layer ASL).

like reaching, grasping, placing, attaching or plugging an object [19]. An observed hand movement that is recognized by the vision system as a particular primitive is represented in the action observation layer (OL). The action simulation layer (ASL) encodes entire chains of action primitives that are in the motor repertoire of the robot (e.g., reaching-grasping-placing/plugging a particular object). These chains are linked to representations of specific goals or end states (e.g., attach right wheel to base) in the intention layer (IL). The basis of the goal inference capacity is the activation of a particular chain and its associated goal during action observation. It is important to stress that due to the self-stabilizing properties of the chain representations, goal inference is possible even if the action sequence performed by the human is only partially observable [11]. The object memory layer (OML) encodes the memory about the position of objects in each of the working areas. The common subgoal layer (CSGL) contains the information about currently active and future subgoals as well as memorized information about subtasks which have been already completed by the team. The construction plan is encoded in the connections between neural populations in 3 different layers representing past, current and future subtasks, respectively. The representations are updated in accordance with the construction plan and real or anticipated feedback from the vision system and/or layer IL. The error monitoring layer (EL) represents a detected discrepancy between the inferred goal of the human partner and the subgoals that are currently available. This error-related activity is functionally relevant since it is linked to representations of compensatory behaviour in the action execution layer (AEL). This layer integrates input from IL, OML, CSGL and EL to select among all possible action sequences the most appropriate complementary sequence. It is worth noting that this layer contains also the representation of an 'action' linked to the speech synthesizer that allows the robot to verbally inform the user about its reasoning.

4 BASIC CONCEPTS OF THE DYNAMIC NEURAL FIELD FRAMEWORK

Each layer of the distributed control architecture is formalized by one or more Dynamic Neural Fields (DNFs). DNFs implement the idea that task-relevant information about action goals, motor primitives or context is encoded by means of activation patterns of local pools of neurons. These patterns are initially triggered by transient input from connected populations and sources external to the network. They may become self-sustained in the absence of any external input due to the recurrent interactions within the population. Functionally, these patterns may thus serve a working memory function. We employed a particular form of a DNF first analyzed by Amari (1977) [2]. In each model layer i, the activity $u_i(x,t)$ at time t of a neuron at field location x is described by the following integro-differential equation (for an overview about analytical results see [8]):

$$\tau_{\mathrm{i}} \frac{\delta u_{\mathrm{i}}(x,t)}{\delta t} = -u_{\mathrm{i}}(x,t) + S_{\mathrm{i}}(x,t) + \int w_{\mathrm{i}}(x-x') f_{\mathrm{i}}(u_{\mathrm{i}}(x',t)) dx' + h_{\mathrm{i}} \quad (1)$$

where the constants $\tau_i > 0$ and $h_i < 0$ define the time scale and the resting level of the field dynamics, respectively. The integral term describes the intra-field interactions. It is assumed 1) that the interaction strength, $w(x,x')$, between any two neurons x and x' depends only on the distance between locations, and 2) that nearby cells excite each other, whereas separated pairs of cells have a mutually inhibitory influence. For the present implementation we used the following integral kernel of lateral-inhibition type:

$$w(x) = A\exp(-x^2/2\sigma^2) - w_{inhib} \quad (2)$$

where $w_{inhib} > 0$ is a constant and $A > 0$ and $\sigma > 0$ describe the amplitude and the standard deviation of a Gaussian, respectively. Only sufficiently activated neurons contribute to interaction. The threshold function $f(u)$ is chosen of sigmoidal shape with slope parameter β and threshold u_0:

$$f_{\mathrm{i}}(u_{\mathrm{i}}) = \frac{1}{1+exp\left(-\beta(u_{\mathrm{i}}-u_0)\right)}. \quad (3)$$

The model parameters are adjusted to guarantee that the field dynamics is bi-stable, that is, the attractor state of a localized activation pattern coexists with a stable homogenous activation distribution that represents the absence of specific information. If the summed input to a local population is sufficiently strong, the homogeneous state loses stability and a localized pattern evolves. Weaker external signals lead to a subthreshold, input-driven activation patten in which the contribution of the interactions is negligible. Normally, a constant input from a single population does not drive directly connected populations. It may play nevertheless an important role for the processing in the joint action circuit. The preshaping by weak input brings populations closer to the threshold for triggering the self-sustaining interactions and thus biases the decision processes linked to behavior. Much like prior distributions in the Bayesian sense, multi-modal patterns of subthreshold activation in for instance the action execution layer (AEL) may represent the probability of different complementary actions [6].

The summed input from connected fields u_{j} is given as $S_{\mathrm{i}}(x,t) = k\sum_{\mathrm{j}} f_{\mathrm{j}}(u_{\mathrm{j}}(x,t))$. The parameter k scales the total input relative to the threshold for triggering a self-sustained pattern.

This guarantees that the inter-field coupling is weak and the field dynamics is dominated by the recurrent interactions. The external inputs from the vision system to layers OL and MOL that initiate the dynamic interplay of the different populations in the network are modeled as Gaussian functions.

The existence of a single self-stabilized pattern of activation in a dynamic field is closely linked to decision making. In layers ASL, IL and AEL subpopulations encoding different chains (ASL), goals (IL) and complementary actions (AEL), respectively, interact through lateral inhibition. This inhibitory interaction leads to the suppression of activity below threshold in competing neural pools whenever a certain subpopulation becomes activated above threshold. To represent and memorize simultaneously 1) the location of several objects of a certain type, and 2) multiple common subgoals, the interaction kernels in layers OML and CSGL were adapted to allow for the existence of multiple patterns of activation. [8]. OML contains individual fields for each of the object classes. They are labeled by the workspace to which the object belongs, that is, each class is represented by two separate fields.

5 RESULTS

In the following we validate the dynamic field architecture by presenting snapshots of the human-robot interaction in the construction task. The examples illustrate the impact of action observation on action selection in varying context from the perspective of the robot. The videos of the human-robot interaction and the associated dynamics of the fields can be found at
http://dei-s1.dei.uminho.pt/pessoas/estela/JASTvideos.htm.
In each case, the robot is supposed to know what common subgoals are currently active and can be selected by the team. In the present implementation real or anticipated feedback about accomplished subtasks trigger directly through hand-coded connections the population representing subsequent assembly steps.

5.1 Anticipatory action selection

The capacity to simulate the consequences of observed actions allows the robot to act in anticipation of the partner's motor intentions. Depending on the predicted outcome of the ongoing action, a sociable robot may for instance decide to already prepare for an action that best serves future needs of the user. Within the dynamic field architecture this is possible since current subgoals of the team are updated based on the inferred goal. The anticipatory action selection is illustrated in Fig. 3. All components are distributed on the table and the user starts the joint construction by grasping a wheel (green object) from above (full grip). The robot has sequences in its motor repertoire that associate the type of grasping with specific goals. A grasping from above is used to attach a wheel to the platform whereas using a side grip is the most comfortable and secure way to hand the wheel over to the teammate. The observation of the full grip (represented in OML) triggers an activation peak in ASL that represents the simulation of the respective reaching-grasping-plugging chain. Since attaching a wheel on the side of the user is a current subgoal for the team, the inputs from layers ASL and CSGL automatically activate the representation of that goal in the intention layer (IL). The existence of this activation pattern initiate a dynamic updating process in layer CSGL (not shown here). The peak representing the subgoal ”attach wheel” disappears and an activation pattern representing the new subgoal ”fix wheel with bolt” evolves. Since all bolts (magenta objects) are in the workspace of the robot, the inputs from layers OML and CSGL converge on a population in the action execution layer (AEL) that represents a decision for a ”hand over bolt” sequence as a complementary behaviour of the robot. As can be seen in the activation pattern of layer AEL, the possible alternative to select a wheel in it's working area with the goal to attach it is also represented by a weaker, subthreshold peak. The decision to serve the user first is the result of small biases in the connection strengths to the populations in CSGL that favor the subtasks and intentions of the user over the subtasks to be realized by the robot. For HRI this offers interesting perspectives since a simple adjustment of these weights will affect how social the robot companion behaves.

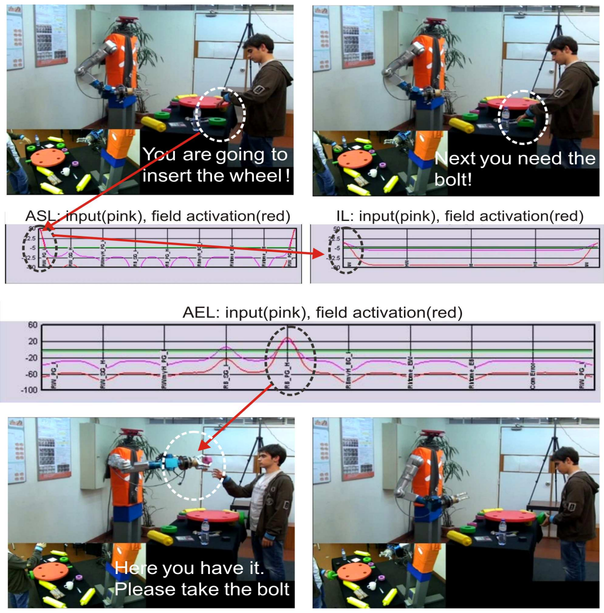


Figure 3. Anticipatory action selection. The human reaches and grasps the wheel from above. The robot infers that the human is going to attach it to the platform. The robot decides to grasp the wheel for handing it over since the wheel is the next component the human will need. The green line in the plots indicates the resting level of the field dynamics.

5.2 Impact of shared task knowledge and context

Very often motor simulation alone is not sufficient to read the motor intentions of the human user. The integration of shared task knowledge is equally important for the decision process [21]. This is illustrated in panel A and B of Fig. 4 In both situations the user reaches his open hand towards the robot. The robot has this gesture which is associated with the goal ”request object” in its motor repertoire, but needs additional information from the common subgoal layer (CSGL) to disambiguate what object the human user is requesting. In panel A the self-stabilized bi-modal activation pattern in CSGL indicates that the two wheels have still to be attached. Since both wheels are located in it's workspace, the robot is able to infer that the user is asking for wheel to attach it (compare the peak in IL). The inputs from layers OML and IL activate a population representation in AEL representing the handing-over sequence. In panel B, the user shows again the requesting gesture. However, the state of the construction and therefor the current subgoals for the team have changed. The bi-modal activation pattern in CSGL now represents the information that the two wheels have to be fixed with bolts. The robot is thus able to infer what object the human wants. A possible complementary action is again to serve the user by handing over a bolt. However, the robot decides instead to attract the attention of the user to the fact that he has a bolt in his working area. The robot performs a pointing gesture in the direction of the bolt. This action selection, which overrides the prepotent tendency to satisfy a user request, is possible because of additional input from the error layer (EL). Population activity in the object memory layer representing the bolt in the workspace of the user together with the input from the population encoding the inferred goal in IL automatically activate a self-stabilized peak in EL. This pattern is associated with the pointing gesture and generates the strongest input to AEL.

5.3 Error detection

The last example shows that even in well known joint action tasks the user can easily make errors that should be compensated by the teammate if possible. Errors may occur for different reasons. The user may have overlooked an object or may be confused about the state of the construction and its temporal order. Different error categories [22] affect joint action on different levels (e.g., error in intention versus error in the selection of action means). The following example illustrates a case in which the robot detect a mismatch between the inferred intention of the user and the state of the construction, that is, between the intention and possible subgoals. In panel A of Fig. 5, the robot observes the human user grasping a wheel from the side which it interprets via action simulation as belonging to a handing over sequence. However, on the side of the robot the wheel is already attached. The information about the already accomplished subtask is memorized by a self-stabilized activation peak in CSGL (compare the snapshot of ”past” field). Input from this field together with input from the intention layer (IL) trigger the emergence of a suprathreshold activation pattern in the error layer (EL). In this case, the error related activity is linked with a population in AEL that initiates speech to explain the nature of the error to the user. The content of the speech combines the information represented in the activation patterns that have initially triggered the error-related activity (”I do not need a wheel since a wheel is already attached on my side”). The example in panel B shows that the information represented in the activation patterns of the various populations of the distributed network can be used to give an even more detailed explanation of a detected error. In this case, the user holds out a bolt to hand it over so that the robot may fix the wheel on its side. Since both wheels have been attached but not yet fixed, the inferred goal is valid. However, the activation pattern in the object memory layer indicates that the robot does not need a bolt since it has one in its working area. Moreover, the user still has to use the bolt himself to continue the assembly of the toy on his side. The activation peak that evolves in EL in response to the converging inputs from layers OML and CSGL controls the speech output. The robot refuses the offered object and informs the user in addition about the missing bolt on his side.

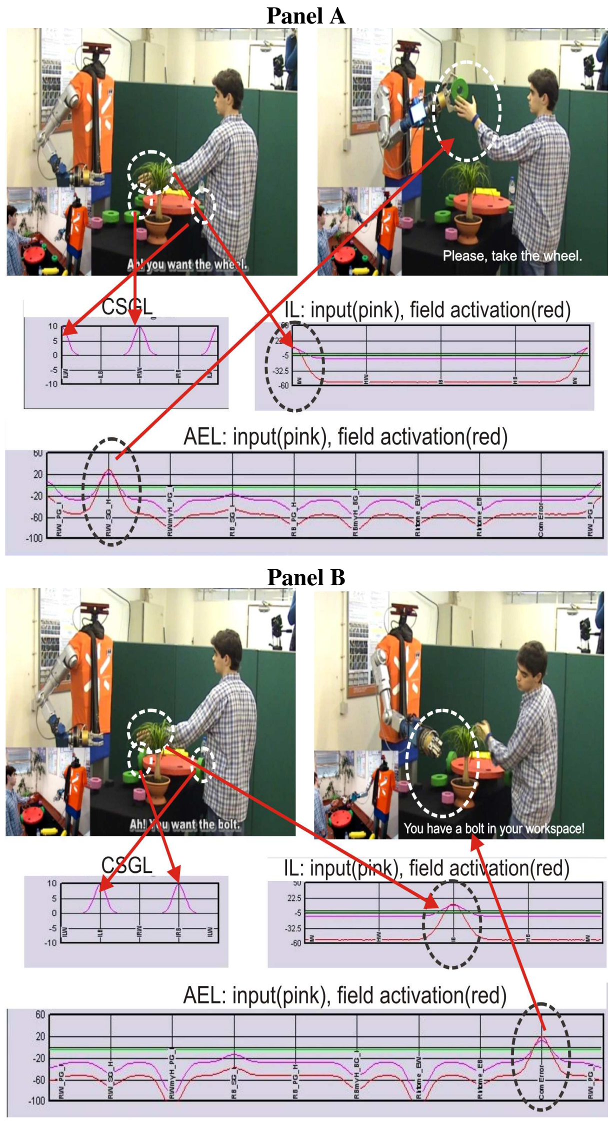


Figure 4. The same observed action may have different meanings. In two different contexts, the human reaches his empty hand toward the robot. Panel A: The robot infers that the human is requesting a wheel with the intention to attach it. The robot decides to grasp it for handing it over. Panel B: The robot infers that the human is asking for a bolt but interprets this request as an error since the human has a bolt in his workspace. The robot decides to communicate this error to the human by pointing to the bolt and speaking to the user.

6 DISCUSSION

The capacity to anticipate and take into account action goals of a partner is considered a fundamental cognitive capacity for successful joint action [21, 18]. We have presented a robot control architecture for human-robot interaction that is based on theories about how humans perceive and act in a social context. The ease with which humans coordinate in routine joint tasks their actions and decisions in space and time is impressive. The capacity to quickly register the intention of the teammate before the action sequence is completed is essential for a fluent team performance. The dynamic field architecture implements the idea that in known tasks dynamic decision making, goal-directed action selection and performance monitoring occur rather effortlessly and do not require a fully developed human capacity for conscious control [16]. As the representation of context, goals and shared task knowledge are interconnected, the observation of a motor act together with situational cues may directly activate the self-sustained population representations of the related goal and the

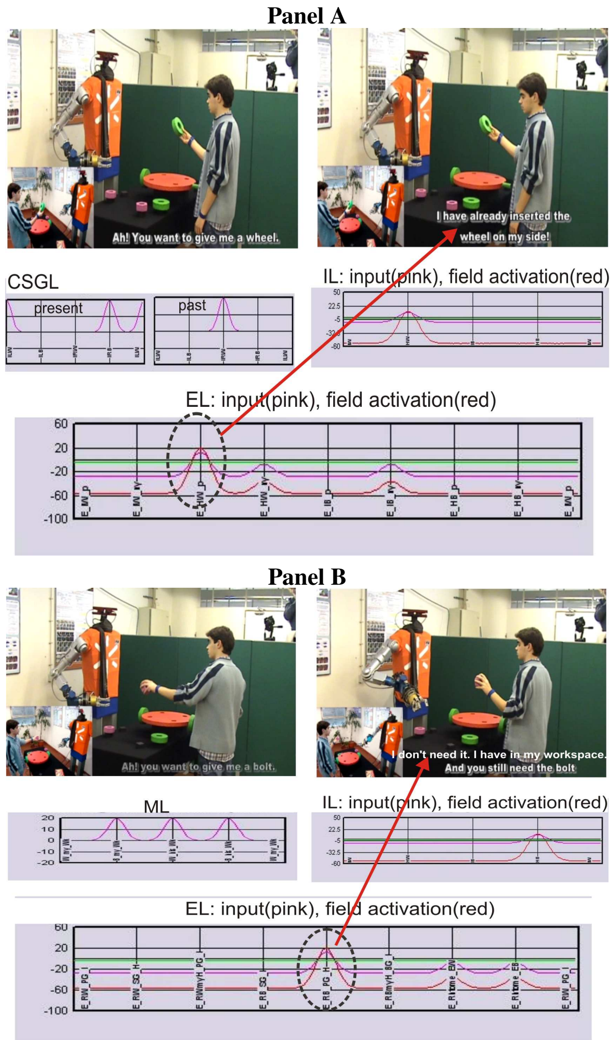


Figure 5. : Error monitoring. Panel A: The robot infers that the human's intention is to hand-over a wheel. The robot interprets this is an error, because the wheel on its side of the platform has been already attached, and decides to communicate this to the human by speaking and nodding its head. Panel B: The robot infers that the human intents to hand over the bolt. The robot interprets the action as an error because the robot has a bolt in its workspace and there is still the need to attach the bolt at the side of the human.

most appropriate complementary action. This automatic process includes basic forms of error monitoring and error compensation

More traditional probabilistic approaches have been applied in the past as well to model and implement cognitive skills like goal inference and decision making for joint action ([7],[17]). Hoffman and Breazeal for instance modeled anticipatory decision making in a Baysian framework to study team fluency in a simulated construction task. In general, Baysian statistics offers a powerful tool for describing human behaviour under circumstances of uncertainty [14]. In our view, a major advantage of the dynamic field approach is that it represents explicitly the important temporal dimension of goal coordination in joint action [21]. Importantly, Dynamic Neural Fields can be used in the Baysian sense by exploiting that multi-modal, subthreshold activation patterns may encode the probability of choices [6]. We are currently testing the joint action model in more complex construction tasks in which the robot has first to infer from observed actions and contextual cues which of several possible toy objects the user is going to built. The accumulated evidence for each of the possible choices is represented by the level of pre-activation of neural populations encoding the different objects.

The decision process linked to complementary actions unfolds over time under multiple influences which are themselves modelled as dynamic representations with proper time scales. This is the basis of flexible behaviour in dynamic joint action conditions. The absence or delay of information about the intention of the user for instance will automatically lead to a decision about an action that does not take into account the other [3]. A challenge for the future will be to endow the robot with the capacity to self-adapt the time window for the integration of input to the dynamics of the different users.

Learning is in general an important research topic of our group. For the present experiments, all inter-field connections were hand-coded. It is certainly not realistic to assume for the next future that for a complex joint action model these connections will self-organize with only modest intervention by the human designer. However, using correlation based learning rules, we have shown in previous work for instance how the goal-directed mappings of the action understanding model may develop during learning and practice [10, 9]. Interestingly, the development process includes the emergence of new task-specific populations which have not been introduced to the architecture by the human designer [11].

The focus of the presented work on action understanding does not mean of course that other information channels of human-human cooperation are not equally important. Our robot is equipped with a speech synthesizer to communicate the state of its reasoning to the user. An obvious extension of this work is to close the loop and advance toward natural-language dialogue. We are currently starting to test a hybrid control architecture that allows us to combine non-verbal and verbal communication skills [13].

ACKNOWLEDGEMENTS

We would like to thank Rui Silva, Eliana Costa e Silva, Toni Machado and Emanuel Sousa for their contribution in many ways.

REFERENCES

[1] A Alissandrakis, C L Nehaniv, and K Dautenhahn, 'Imitation with alice: learning to imitate corresponding actions across dissimiliar embodiments.', *IEEE Transactions on Systems, Man and Cybernetics-PartA*, **32**, 482–496, (2002).

[2] S Amari, 'Dynamics of pattern formation in lateral-inhibitory type neural fields.', *Biological Cybernetics*, **27**, 77–87, (1977).

[3] E Bicho, L Louro, N Hiplito, and W Erlhagen, 'A dynamic neural field architecture for flexible and fluent human-robot interaction', in *Proceedings of the 2008 International Conference on Cognitive Systems*, pp. 179–185. University of Karlsruhe, Germany, (2008).

[4] C Breazeal, 'Social intercations in HRI: The robot view', *IEEE Transactions on Systems, Man and Cybernetics-PartC: Applications and Reviews*, **34**, 181–186, (2004).

[5] S Calinon, F Guenter, and A Billard, 'On learning, representing and generalizing a task in a humanoid robot.', *IEEE Transactions onSystemsm man and Cybernetics, Part B*, **37**, 286–298, (2007).

[6] R Cuijpers and W Erlhagen, 'Implementing Bayes' rules with neural fields', in *ICANN 2008, Part II, LNCS 5164*, ed., V Kurkov, pp. 228–237. Springer Verlag, (2008).

[7] R H Cuijpers, H T van Schie, M Koppen, W Erlhagen, and H Bekkering, 'Goals and means in action observation: A computational appraoch', *Neural Networks*, **19**, 311–322, (2006).

[8] W Erlhagen and E Bicho, 'The dynamic neural field approach to cognitive robotics', *Journal of Neural Enginering*, **3**, R36–R54, (2006).

[9] W Erlhagen, A Mukovskiy, and E Bicho, 'A dynamic model for action understanding and goal-directed imitation.', *Brain Research*, **1083**, 174–188, (2006).

[10] W Erlhagen, A Mukovskiy, E Bicho, G Panin, C Kiss, A Knoll, H van Schie, and H Bekkering, 'Goal-directed imitation for robots: a bioinspired approach to action understanding and skill learning', *Robotics and Autonomous Systems*, **54**, 353–360, (2006).

[11] W. Erlhagen, A. Mukovskiy, F. Chersi, and E. Bicho, 'On the development of intention understanding for joint action tasks.', in *6th IEEE Int. Conf. on Development and Learning*, pp. 140–145. Imperial College London, (11-13 July, 2007).

[12] T Fong, I Nourbakhsh, and K Dautenhahn, 'A survey of socially interactive robots', *Robotics and Autonomus Systems*, **42**, 143–166, (2003).

[13] M E Foster, M Giuliani, T Mller, M Rickert, A Knoll, W Erlhagen, E Bicho, N Hiplito, and L Louro, 'Combining Goal Inference and Natural-Language Dialogue for Human-Robot Joint Action', in *Proceedings of the 1st International Workshop on Combinations of Intelligent Methods and Applications (CIMA 2008)*. Patras, Greece, (July, 2008).

[14] J I Gold and M N Shadlen, 'Banburismus and the brain: the relationship between sensory stimuli, decisions, and reward.', *Neuron*, **36**, 299–308, (2002).

[15] S Gray, C Breazeal, M Berlin, A Brooks, and J Lieberman, 'Action parsing and goal inference using self as simulator', *IEEE International Workshop on Robots and Human Interacive Communication*, 202–209, (2005).

[16] R R Hassin, H Aarts, and M J Ferguson, 'Automatic goal inferences', *Journal of Experimental Social Psychology*, **41**, 129–140, (2005).

[17] G Hoffman and C Breazeal, 'Cost-based anticipatory action selection for human-robot fluency', *IEEE Transactions on Robotics*, **23**, 952–961, (2007).

[18] R D Newman-Norlund, M L Noordzij, R G J Meulenbroek, and H Bekkering, 'Exploring the basis of joint action: Coordination of actions, goals and intentions', *Social Neuroscience*, **2**, 48–65, (2007).

[19] S Schaal, 'Is imitation learning the route to humanoid robots?', *Trends in Cognitive Science*, **3**, 233–242, (1999).

[20] G Schöner, 'Dynamical systems approaches to cognition', in *The Cambridge Handbook of Computational Psychology*, ed., R Sun, pp. 101–125. Cambridge University Press, (2008).

[21] N Sebanz, H Bekkering, and G Knoblich, 'Joint action: bodies and minds moving together', *Trends in Cognitive Sciences*, **10**, 70–76, (2006).

[22] T P Spexard, M Hanheide, S Li, and Wrede B, 'Oops, something is wrong: Error detection and recovery for advanced human-robot intercatios', in *Proceedimgs of the ICRA 2008 Workshop on Social Intercation with Intelligent Indoor Robots*, eds., G J Kruijff, H Zender, M Hanheide, and B Wrede, (2008).

[23] H T van Schie, B M van Waterschoot, and H Bekkering, 'Understanding action beyond imitation: Reversed compatibilty effects of action observation in imitation and joint action', *Jurnal fo Experimental Psychology: Human Perception and Performance*, (in press).

[24] G. Westphal, C. von der Malsburg, and R. P. Würtz, *Applied Pattern Recognition*, chapter Feature-Driven Emergence of Model Graphs for Object Recognition and Categorization, 155–199, Springer Verlag, 2008.

[25] M Wilson and G Knoblich, 'The case for motor involvment in perceiving conspecifics', *Psychological Bulletin*, **131**, 460–473, (2005).

Using Visual Attention to Evaluate Collaborative Control Architectures for Human Robot Interaction

Tom Carlson[1] and **Yiannis Demiris**[2]

Abstract. Collaborative control architectures assist human users in performing tasks, without undermining their capabilities or curtailing the natural development of their skills. In this study, we evaluate our collaborative control architecture by investigating the visual attention patterns of robotic wheelchair users. Our initial hypothesis stated that the user would require less visual attention for driving, whilst they are being assisted by the collaborative system, thus allowing them to concentrate on higher level cognitive tasks, such as planning. However, our analysis of eye gaze patterns—as recorded by a head mounted eye tracking system—supports the opposite conclusion: that patterns of saccadic activation increase and become more chaotic under the assisted mode. Our findings highlight the necessity for techniques that assist the user in forming an appropriate mental model of the collaborative control architecture.

1 INTRODUCTION

Smart wheelchairs are being developed to augment a mobility impaired person's capabilities, enabling them to safely perform precise manoeuvres. In order to achieve this, we must share the control appropriately between the user and the robotic chair, such that the user still feels in control [10]. The human driver knows what they want to achieve and is good at interpreting complex, cluttered environments, however a robot can be much more precise in executing low level commands. Therefore, we have proposed an effective collaborative control methodology, which infers the user's intentions from their joystick input, along with the affordances of the surrounding environment [3, 1]. Based on these predictions, the wheelchair can alter the motor control signals to provide assistance, where necessary.

Our collaborative controller successfully increased the safety of trajectories driven in narrow spaces, whilst simultaneously reducing the need for excessive corrective joystick movements [2]. However, after processing feedback from these earlier trials, we are now investigating the possibility of using additional physiological input from the user, to help the wheelchair behave as naturally as possible during interaction. We propose to utilise the user's eye gaze, to estimate their attention, which could simultaneously enhance our prediction of intention algorithm and indicate when the wheelchair does not behave as the user expects.

Whilst the user is being assisted by the collaborative system, we hypothesize that the user would require less visual attention to effectively manoeuvre the wheelchair. This would allow them to simultaneously perform other higher level cognitive tasks, such as environmental exploration, or planning future manoeuvres. We also expect the driver to fixate on objects of interest, which may help to strengthen our intent-prediction system.

We do not treat eye gaze as an active input device, in which the user tries to control the wheelchair by moving their head and/or eyes, as was demonstrated in [9]. Instead, we aim to use it as a passive device, to non-intrusively increase the user state vector (the knowledge we possess about the user at each time step).

In this exploratory study, we observe the characteristics of the user's eye movements, whilst performing typical manoeuvres, such as driving around offices and passing through narrow doorways. The observations are made over one independent variable, which can take one of two states: provide adaptive assistance, or provide no assistance.

Figure 1. A participant driving our robotic wheelchair. The software on the tablet PC combines the stimulus from the joystick with the localisation data derived from the camera, to collaborate with the user in controlling the wheelchair motion. Simultaneously, the head-mounted eye-tracking system records the driver's gaze.

2 BACKGROUND

In this section, we will introduce the robotic wheelchair platform that we have developed (Figure 1). After briefly describing the hardware

[1] Imperial College London, UK, email: tom.carlson02@imperial.ac.uk, web: http://www.iis.ee.ic.ac.uk/~tcarlson

[2] Imperial College London, UK, email: y.demiris@imperial.ac.uk, web: http://www.iis.ee.ic.ac.uk/yiannis

components, we focus more on the underlying collaborative control architecture. We then go on to explain the eye tracking system, describing the hardware setup along with the fundamental operation of the tracking algorithm.

2.1 The Robotic Wheelchair Platform

Our system is built upon a mid-wheel drive EPIOC (electrically powered indoor/outdoor chair), that would typically be prescribed to a severely mobility impaired patient. We have interfaced a tablet PC with the wheelchair's joystick and motor control unit. The joystick signals are intercepted and altered, where necessary, before being forwarded to the wheelchair's motor control unit.

The collaborative controller (Figure 2) is realised in software on the tablet PC. The adaptive assistance is provided by the shared controller module, which uses information from the *safe mini-trajectory* generator, along with the intention predictor, to decide exactly how to adapt the joystick signals [3]. The system is currently underpinned by the computer vision-based localisation system that we have developed to work in mapped, indoor environments (with minimal modification of the environment) [3].

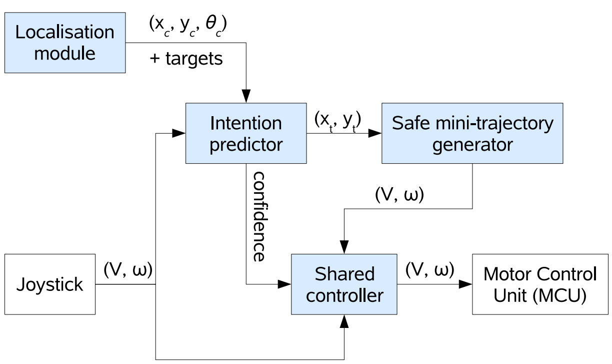


Figure 2. The collaborative system shares the control appropriately between the user and on-board computer [3]. (x_c, y_c, θ_c) and (x_t, y_t) describe the wheelchair's current and target positions respectively. (V, ω) represent the target translational and rotational velocity tuple to be sent to the motor control unit.

2.1.1 *Intention prediction*

In our system, we base our intention prediction on the multiple hypothesis approach, as described in [4]. Our prediction models are task based, so we define targets of interest, such as doorways and desks, which the user may wish to drive through or approach. We constructed a confidence function (Equation 1), which only increases when moving towards a target. This function is the product of two parts: the first (Equation 2) is computed using the Euclidean distance from the current wheelchair position $\mathbf{w}$ to the target $\mathbf{w_t}$, the second (Equation 4) is based upon the heading of the chair θ, compared with the angle to the target ϕ (Equation 3). The scaling factor of k in Equation 4 determines the sensitivity towards the angular error and in our case was experimentally set to 2.

$$C = C_d C_\theta \tag{1}$$

$$C_d = \exp\left\{-||\mathbf{w_t} - \mathbf{w}||\right\} \tag{2}$$

$$\phi = \arg\left(\mathbf{w_t} - \mathbf{w}\right) \tag{3}$$

$$C_\theta = \exp\left\{\frac{k(\pi - |\theta - \phi|)}{\pi} - k\right\} \tag{4}$$

Essentially the *safe mini-trajectory* generator computes a path to reach the predicted target safely, once the confidence threshold has been reached for that target. The wheelchair is then guided gently towards the first waypoint of the *safe* path. However, we allow the user to gradually deviate from this path, if they consistently oppose this attraction. The confidence value will then fall accordingly; eventually allowing them to regain full control if appropriate. Conversely, we will prevent them from deviating from the safe path if it is going to result in a collision (e.g. they are in a doorway and might hit the doorframe). However, the speed of the manoeuvre is always controlled by the user, in a manner similar to that of Zeng et al. [12]; it is proportional to the component of the joystick input vector, which lies in the *safe* direction. This continues until the corresponding confidence value has dropped below an experimentally set threshold, C_{thresh}, which occurs once the chair has safely reached its target destination. We also allow the user to reverse backwards along the safe path at any time, until the confidence value drops below C_{thresh}, at which point they regain full control. This strategy strives to make the user feel more in control than using a rigid assistance method, which forces them to stay on a computer-controlled path at all times, for example, when the *CALL smart wheelchair* uses its "optical track follower" [10].

2.2 The Eye Tracker

In the BioART lab, we have constructed a portable monocular eye-tracker, which is based on the openEyes system developed at Iowa State University [7]. It allows us to indicate the user's point of gaze (POG) on a projection of their field of view (the scene image) [11]. We decided to use a cycle helmet as our substrate, to comfortably and securely support the hardware, whilst allowing quick adjustments to be made for new users (Figure 5).

The scene image is produced by a firewire camera with a fish-eye lens (111° field of view), mounted on the headpiece above the tracked eye, which reduces parallax error. Concurrently, the subject's right eye is illuminated by an infra-red LED, which is observed by a second firewire camera. To reduce the sensitivity to lighting conditions, this second camera is fitted with a Kodak Wratten 87C infrared filter, which blocks the visible spectrum.

2.2.1 *The PCCR algorithm*

The eye-tracking algorithm uses the pupil centre and corneal reflection method (PCCR) [5]. We based our implementation on the openEyes project [7], which utilises the Starburst algorithm [8]. Our implementation is in C++ rather than C, which has allowed us to make a portable class-based solution, abstracting away the interface dependence on the HighGUI/GTK widgets. This means we can easily use the real-time eye tracking class in our existing multithreaded QT framework, or as a standalone application.

Since we are using filtered, reflected infrared light, the corneal reflection should be the brightest part of the image, as shown in Figure 3(a). An adaptive binary thresholding method is used to detect this. Once the reflection has been found, its position is recorded and it is subsequently removed from the image, so as not to interfere with the pupil detection. This is done by replacing it with the average pixel intensity of its neighbours.

(a) Corneal reflection. (b) Finding the pupil.

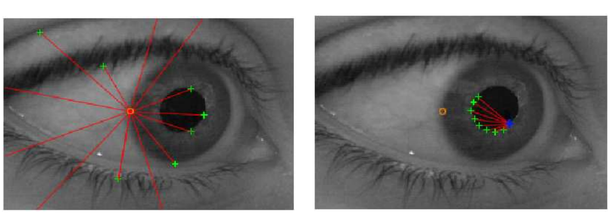
Figure 3. The gaze direction is computed from the vector between the centre of the pupil and the corneal reflection. To do this, we use a modified version of Starburst algorithm [8].

We have extended this method to prevent erroneous detections that sometimes occur in bright lighting conditions, or if the user is wearing glasses. Once the pupil has been detected, we check that the corneal reflection lies within 150 pixels of the pupil centre. If this is not the case, we discard the reflection result and restart the search, looking for the next largest area of high intensity pixels.

2.2.2 *Determining the pupil centre*

Robustly finding the pupil is more of a challenge, we apply the feature-point detection method which is also part of the Starburst algorithm [8]. We start by setting the estimate of the centre of the pupil $\mathbf{p_{current}}$ to be in the centre of the image, at each iteration, this estimate is updated to be the mean of the generated feature set. In successive frames, $\mathbf{p_{current}}$ is taken to be the final value from the previous image.

The first part of the algorithm generates a number of rays radiating from $\mathbf{p_{current}}$, to the edges of the image. Each ray is then traversed, calculating the derivative intensity between neighbouring pixels, Δ. For the first point $\mathbf{f_i}$ on each ray that this intensity is greater than a certain threshold, δ, it is deemed to be a potential boundary between the pupil and iris. Consequently it is added to the candidate feature set (Figure 4(a)). Once all the rays have been traversed, a similar process is begun for each of the points in the generated feature set, only this time, the rays are constrained to fall within an arbitrary angle of $\mathbf{p_{current}}$ (Figure 4(b)). The whole algorithm is repeated with $\mathbf{p_{current}}$ set to the geometric centre of the feature set, unless they are within 10 pixels of each other, in which case it terminates [8].



(a) Phase 1. (b) Phase 2 (1st iteration).)

Figure 4. Detecting the pupil using the Starburst algorithm [8]. In phase 1, potential features (green crosses) are found along n rays radiating from $\mathbf{p_{current}}$ (red blob in centre). In phase 2, new feature points are located along n rays from each of the features $\mathbf{f_i}$ in our current set, back towards $\mathbf{p_{current}}$.

The final stage is to fit an ellipse to the candidate feature set, which is performed using the random sample consensus (RANSAC) paradigm in a similar manner to [6]. This allows the true centre of the pupil to be estimated.

2.2.3 *Calibration*

Now that we have a system for estimating the *eye difference vector* from the centre of the pupil to the corneal reflection, we can determine the POG on the scene image. This is achieved by using the homographic mapping function, H, which is a 3×3 matrix with eight degrees of freedom. H is computed by performing a standard calibration, whereby the user sequentially fixates on 3×3 known grid points in the scene image [8].

Figure 5. Our portable eye-tracker consisits of two firewire cameras; one of which observes the subject's right eye, whilst the other records the scene (the person's field of view). After performing a manual calibration, the software will mark the person's focal point on the scene image.

It is important to note that this eye tracking system does not give pixel-level resolution on the scene image, as might be required for the manipulation of a cursor on a screen. Instead, it is sufficiently accurate to detect when the subject's POG dwells on regions of interest, such as doorways, desks, people etc., however it remains extremely sensitive to any calibration error.

2.2.4 *Improved calibration for inexperienced users*

In our experiments, many of the subjects have never experienced the use of an eye tracking system before, so we must ensure the process is as simple and non-intrusive as possible. Although the calibration procedure is relatively quick and simple, it is crucial to get it right first time. It has proven to be disruptive when running trials with several participants that each have to make successive re-calibrations. We observed that many of these errors occurred due to the subject blinking, or moving the eye just as the calibration point was set.

Therefore, we improved the calibration stage in order to automatically detect and eliminate errors in producing the homographic mapping matrix. We now compute the stability of the *eye difference vector* over a few frames before and after the calibration point has been set. If the vector is deemed unstable (± 5 pixels on each end point), the point is rejected and the subject must perform that fixation again, before moving onto the next grid point. This intuitive adjustment to the algorithm greatly reduced the need for re-calibration, which has saved experimental time, reduced erroneous results and helped to instill greater confidence in our test subjects.

3 METHODOLOGY

This paper presents the approach we followed for conducting a pilot study. It is important to test the methodology before carrying out the full scale trials, which will take a considerable length of time. We followed some generally accepted principles for our experimental design, to ensure our results are as meaningful as possible and eliminate any bias. A script was used to ensure that all participants were told the same information. Additionally, none of them were aware of the goals of the experiment.

In contrast with previous work [2, 3], the participants were not informed prior to the trial whether the system would be assisting them to perform the manoeuvre or not. This allowed us to gain some interesting results relating to people's mental models of the underlying control system, which will be further explored in the Discussion section.

Each participant was asked to drive from the start point in one office, through a narrow doorway to an adjoining office, exit the office via another doorway and stop in the corridor, as shown in Figure 6. This route was driven four times by each participant. Half of the trials were driven with no assistance, whereas the other half were driven with the collaborative controller active, thereby providing assistance at certain points, notably when driving through the doorways. In order to prevent biases, the collaborative controller was active during the odd trials for the odd–numbered participants and during the even trials for the even–numbered participants.

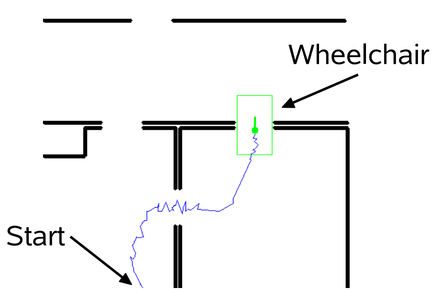


Figure 6. In the trials, participants were required to drive from the start position in one office through to an adjoining office, before arriving at the finish position in the corridor. When using the collaborative controller, it provides assistance, where necessary, in the approach to and whilst driving through the doorways.

In our previous experiments [3, 2], we received many useful comments from the participants, so this time we decided to collect information that we could quantify in a structured questionnaire. After driving each trajectory, the participant was asked to rate the following carefully chosen statements on a five point Likert scale, ranging from "strongly agree" to "strongly disagree":

- The wheelchair was easy to manoeuvre.
- The wheelchair behaved as I expected.
- I had to concentrate hard to drive the wheelchair.
- It felt natural driving the wheelchair.

4 RESULTS

We took two approaches to processing the large quantity of video data that we had collected. The first was to qualitatively gain information from watching the playback of the scene images, which had been automatically annotated with the gaze points. The second was to perform a quantitative analysis, by computing statistics on clusters of the gaze points, for instance: their standard deviation, whilst passing through a doorway.

Using our first approach, we noticed that all the participants tended to perform a visual exploration of the scene as they drove through through a doorway, an example of which is shown in Figure 7. This comprised of rapid saccadic eye movements. However, once they were in a room they tended to fixate on objects or perform smooth pursuit, as shown in Figure 8.

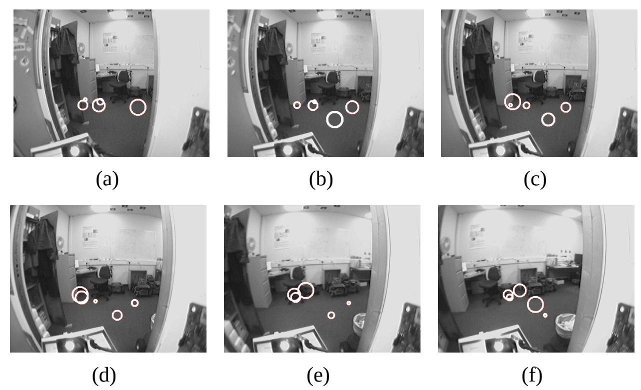


Figure 7. A sequence showing the visual exploration as a user drives through a doorway. The white circles indicate the gaze of the user; the largest one is the current POG, the smaller ones indicate the previous four POGs.

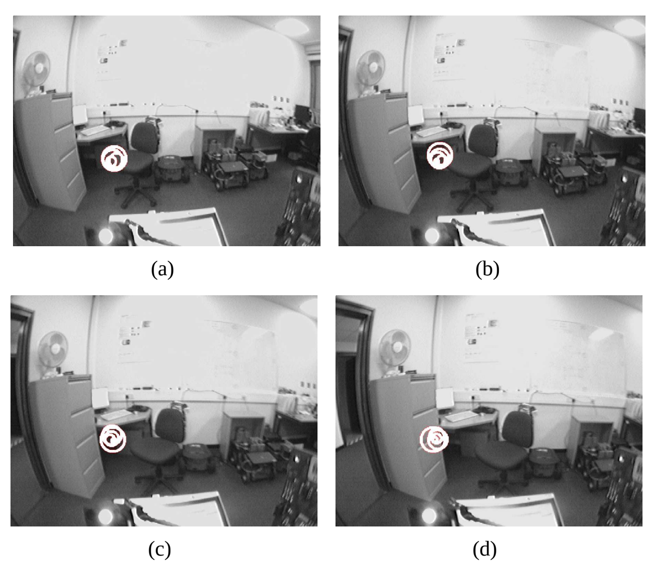


Figure 8. Once a user is in an office, which they have already visually explored, their gaze tends towards the forward position. This is clearly shown by the sequence of processed POGs, which form approximate concentric circles.

Typically, users driving without any assistance exhibited patterns of movement in the points of gaze (POGs) as shown in Figure 9. In this case, the driver was looking predominantly straight ahead, in the direction of the chair's movement. However, when the collaborative control system assisted the user (unbeknown to them) to drive through a narrow doorway, the typical pattern of POGs changed dramatically to resemble that of Figure 10.

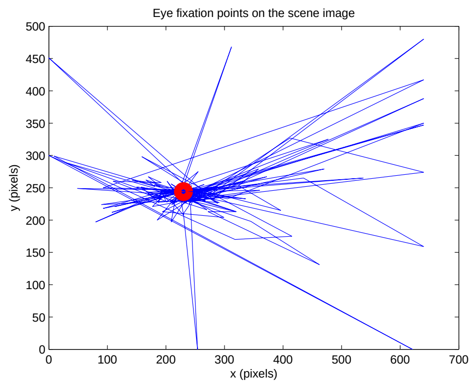


Figure 9. A superposition of the the typical POGs, when driving through a doorway without assistance. The large red blob indicates the median, which corresponds to the user looking straight ahead.

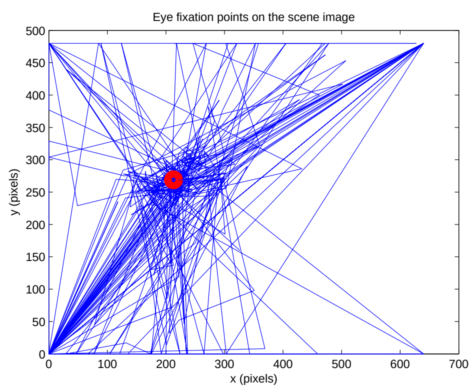
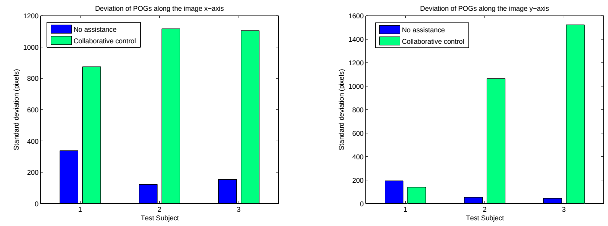


Figure 10. A superposition of the the typical POGs, when driving through a doorway using the collaborative control method. The large red blob indicates the median, which corresponds to the user looking straight ahead.

5 DISCUSSION

When people were driving without any assistance, their POGs predominantly clustered around the median position, with a relatively low deviation, as you can see in Figure 9. The median position corresponds to the times at which the user is looking straight ahead. In Figure 11, the graphs highlight the significant increase in the deviation of the POGs, once the user is being assisted by the collaborative controller. All the subjects exhibited increased deviation in the horizontal plane of the scene image (Figure 11(a)), however the results for the vertical plane of the image (Figure 11(b)) may yield an interesting explantion for the increased saccadic eye movements.



(a) Scene x-axis.

(b) Scene y-axis.

Figure 11. The standard deviation of the points of gaze (POGs) for subjects whilst driving through narrow doorways. The case when no assistance was given is compared with the case when the collaborative controller was active.

When using the collaborative controller—compared with not being given any assistance—all of the participants agreed more strongly with the statement: "I had to concentrate hard to drive the wheelchair" and tended to disagree more with "the wheelchair behaved as I expected". However their reasons for this differed. Predominantly people tended to comment about the wheelchair "not behaving correctly", whereas the first participant described feeling that it was their own fault for not understanding how to operate the wheelchair properly. This could explain why the first participant was the only person not to also significantly increase deviation in the vertical plane of the image (Figure 11(b)). They tended to look only at the doorframes (Figure 12), whereas everyone else additionally focussed on the tablet PC.

This suggests a potential lack of a mental model for the wheelchair. We define a person's mental model to be their perceived forward model of a system's behaviour. For example, if we apply a control signal to the current system state, what would be the next state of the system [4]? In our case, how does the user expect the wheelchair to move as a result of a joystick manipulation?

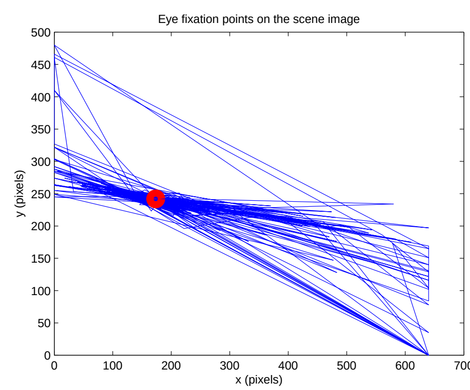
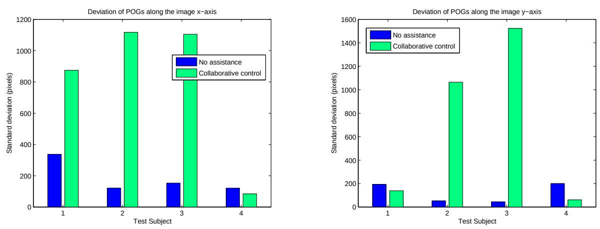


Figure 12. An increased deviation in the horizontal plane, without a significant increase in the vertical plane, suggests the user is more concerned with performing the manoeuvre, than what the control system on the tablet PC is doing.

Inspired by our previous results, we decided to carry out an additional trial, to test our hypothesis regarding the requirement of an adequately formed mental model. The results of which are plotted against our previous findings in Figure 13. We informed an additional fourth test subject about the shared control policy within the collaborative control architecture and how it would assist with manoeuvres whilst it was active. This meant they were able to form an appropriate mental model of the wheelchair's expected dynamic behaviour. Consequently they produced no significant difference in either axis of their eye gaze patterns between the case when they were given assistance and the case when they were in full control.

This would suggest that a high degree of saccadic eye movement could indeed be triggered by the lack of an appropriate mental model. However it would require further trials, which we are currently undertaking in the BioART lab, to produce statistically significant results, to support this premise.



(a) Scene x-axis. (b) Scene y-axis.

Figure 13. The additional fourth test subject was aware of the shared control policy within the collaborative control architecture. This meant they had an appropriate mental model of the wheelchair's behaviour, consequently resulting in no significant difference in the eye gaze patterns.

6 CONCLUSIONS

We have constructed a head-mounted, portable eye-tracking system and interfaced it with our existing robotic wheelchair. Through the use of our collaborative control architecture, people have improved the quality of their driving, in terms of the safety of the trajectories followed. Conversely, an analysis of the user's eye movement, combined with a questionnaire and verbal feedback, has indicated potential difficulties that users have in the recognition and understanding of adaptive interfaces.

In our work, we have again demonstrated the importance of using physiological measures in addition to the more traditional system performance metrics (e.g. speed, distance etc.) when evaluating intelligent HRI architectures. The results can be counter-intuitive; for example, it may require more concentration to perform a task when being given adaptive assistance. However, such results are important to discuss and in this case, they follow the premise that it could take longer to create a mental model of an adaptive interface than of one that has a fixed response.

Our initial hypothesis stated that the user would require less visual attention for driving, whilst they are being assisted by the collaborative system. This would allow them to concentrate on higher level cognitive tasks, such as planning or performing a visual search. However, our analysis of eye gaze patterns for untrained users supports the opposite conclusion; that patterns of saccadic activation increase and become more chaotic under the assisted mode. Therefore, our findings reiterate the necessity for techniques that assist the user in forming an appropriate mental model of the collaborative control architecture.

7 FUTURE WORK

Inspired by our unexpected findings and the success of our pilot study, we intend to perform a full-scale study, investigating the user's mental model of the collaborative control system. It would be interesting to do a *between* subjects trial, where some users only operate the chair with the collaborative controller active, whilst others are not given any assistance. Over the course of a number of trials, we would expect both the groups to form different perceptual models of the wheelchair's behaviour. Despite these differences, the degree of saccadic eye movements may converge.

ACKNOWLEDGEMENTS

The authors would like to thank Kaveh Yousefi, Philipp Leskovitz and Valerio Chang for their contributions to the development of the eye-tracker. We are also extremely grateful to all those who participated in our experiments. Finally, we would like to thank the members of the BioART team — Simon Butler, Murilo Fernandes Martins, Bálint Takács, Paschalis Veskos and Yan Wu — for their insightful comments and continued support.

REFERENCES

[1] Tom Carlson and Yiannis Demiris, 'Evidence of shared control improving accuracy in human-wheelchair interaction', in *Proc. The MAIA Project Workshop 2007*, K.U. Leuven, Belgium, (Nov 2007).

[2] Tom Carlson and Yiannis Demiris, 'Collaborative control in human wheelchair interaction reduces the need for dexterity in precise manoeuvres', in *Robotic Helpers: User Interaction, Interfaces and Companions in Assistive and Therapy Robotics, a Workshop at ACM/IEEE HRI 2008*, pp. 59–66, Amsterdam, The Netherlands, (March 2008).

[3] Tom Carlson and Yiannis Demiris, 'Human-wheelchair collaboration through prediction of intention and adaptive assistance', in *Proc. of IEEE International Conference on Robotics and Automation*, pp. 3926–3931, Pasadena, CA, (2008).

[4] Yiannis Demiris, 'Prediction of intent in robotics and multi-agent systems', *Cognitive Processing*, **8**(3), (September 2007).

[5] E. D. Guestrin and M. Eizenman, 'General theory of remote gaze estimation using the pupil center and corneal reflections', *IEEE Transactions on Biomedical Engineering*, **53**(6), 1124–1133, (June 2006).

[6] Dan Witzner Hansen and John Paulin Hansen, 'Robustifying eye interaction', in *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06),*, pp. 152–159, (2006).

[7] D. Li, J Babcock, and D. J. Parkhurst, 'openeyes: A low-cost head-mounted eye-tracking solution', in *ACM Eye Tracking Research and Applications Symposium*, (2006).

[8] D. Li, D. Winfield, and D. J. Parkhurst, 'Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches', in *IEEE Vision for Human-Computer Interaction Workshop at CVPR*, pp. 1–8, (2005).

[9] Y. Matsumoto, T. Ino, and T. Ogasawara, 'Development of intelligent wheelchair system with face and gaze based interface', in *Proc. of 10th IEEE Int. Workshop on Robot and Human Communication (ROMAN 2001)*, pp. 262–267, (2001).

[10] P. Nisbet, 'Who's intelligent? Wheelchair, driver or both?', in *Proc. IEEE Intl. Conference on Control Applications*, Glasgow, Scotland, U.K., (September 2002).

[11] Kaveh Yousefi, *Multimodal human-computer interaction : eye tracking*, Master's thesis, Department of Computing, Imperial College London, 2007.

[12] Qiang Zeng, Etienne Burdet, Brice Rebsamen, and Chee Leong Teo, 'Evaluation of the collaborative wheelchair assistant system', in *IEEE Conference on Rehabilitation Robotics*, The Netherlands, (June 2007).

Toward Therapist-in-the-Loop Assistive Robotics for Children with Autism and Specific Language Impairment

Mark B. Colton[1], Daniel J. Ricks[1], Michael A. Goodrich[2], Behzad Dariush[3], Kikuo Fujimura[3], and Martin Fujiki[4]

Abstract. In this paper we summarize some of the current work in the field of robot-based therapy for children with autism, and explore new directions that emphasize the important role of the therapist in achieving clinical benefit for these children and children with specific language impairment (SLI). Our ongoing and future research related to robot-based imitation therapy, design of therapist interfaces, and development of therapy robots is described. The overarching assumption in each of these research thrusts is that clinical benefits to children with autism and SLI are best achieved through triadic interactions between the child, the therapist, and the robot facilitator.

1 INTRODUCTION

There is growing anecdotal evidence that robots provide unique opportunities for assisting children with autism. Children with autism exhibit social behaviors with robots that may be useful in generating potential therapies. Such social behaviors include imitation, eye gaze, and joint attention [12,15]. These social behaviors are often rare in children with autism, but evidence suggests that robots trigger them more often in such children; sometimes these behaviors can be prompted and sometimes they are spontaneous.

Unfortunately, there is no evidence that the social behaviors observed in interactions between children with autism and robots will generalize from laboratory or clinical settings to interactions between children with autism and other people. However, it is believed that developing even basic social skills for children with autism will not only allow them to becoming higher functioning, but also enable and reward their caregivers in the challenging task of providing their care.

In addition to children with autism, there is another large population of children that can potentially benefit from using robots to improve social skills. It has been observed that children with specific language impairment (SLI) often lag behind their peers in social development [2,4]. Although there is much debate as to the cause of social deficiencies in children with SLI, there appears to be general agreement that early intervention to promote social skill development will allow these children to have fewer social problems in the near-term and to achieve higher levels in the long-term. Thus, it is desirable to consider how assistive robotics can benefit children on a spectrum of abilities, from low-functioning children with autism to higher-functioning children with SLI.

The purpose of this abstract is to summarize general trends in robot-assisted interactions with children with autism, and outline a research agenda for developing effective therapies for helping not only those children but also children with SLI. We are interested in using robots to help children with SLI, although there appears to be no work on using robots to help this portion of the population even though the number of children with SLI is much larger than the number of children with autism [27].

[1] Dept. of Mechanical Engineering, Brigham Young University, USA. Email: {colton,danricks}@byu.edu.
[2] Dept. of Computer Science, Brigham Young University, USA. Email: mike@cs.byu.edu.
[3] Honda Research Institute, USA. Email: {dariush,kfujimura}@hra.com.
[4] Dept. of Communication Disorders, Brigham Young University, USA. Email: martin fujiki@byu.edu.

2 THERAPIST-IN-THE-LOOP INTERACTION

2.1 Brief Overview of Related Literature

Children with SLI tend to exhibit reticence and solitary-passive withdrawal more than typically developing peers, and children with more severe impairments demonstrate fewer prosocial behaviors [1]. Additionally, emotion regulation and language impairment are significant predictors of reticence [2]. Studies considering why some children with SLI can struggle with social competence point to difficulty in identifying and describing emotional state from a story context [3], as well as in inferring emotion from prosody in vocally-presented stories [4]. Complementing these studies is evidence that children with SLI may also display deficiencies in properly displaying emotions, specifically knowing when to display and when to dissemble emotions in social situations [5].

Robotic technologies provide potential therapies for children with SLI and autism, but it is important that we do not lose sight of the goal of helping children just so we can promote a robotic research agenda. However, there are a number of reasons to consider robots as therapeutic tools for children with autism and SLI. Robots have been shown to invoke interest and social interaction from young children [10] and to engage children with autism at various age levels [11-13]. One of the reasons for using robots in therapies for autistic children is because of the toy-like engagement that it offers [14]. Additionally, work has been performed that uses robots as a tool for diagnosing autism [15]. However, it is important to note that using robots to treat children with autism or SLI should avoid introducing new or reinforcing old stereotypical behaviors in such children [16], including providing an object on which the child may fixate, and should consider including the robot not only as a "social other" but also as a mediator/facilitator for a therapist, clinician, parent, or other child [11, 12]. Moreover, attention must be given to the appearance of the robot because it may be that animal-like robots or simple humanoid forms are more useful than realistic humanoid robots [11, 17]. Finally, it should be noted that robots

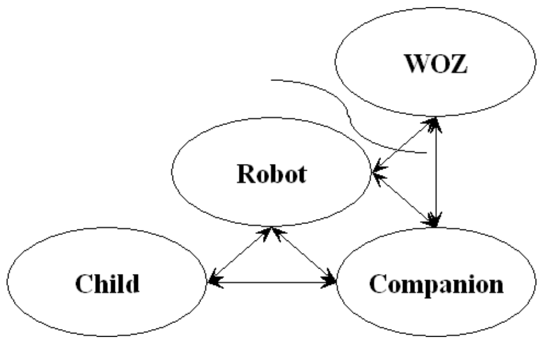


Figure 1. Triadic interactions between child, robot, and companion.

are not the only computer-based technologies that are useful in autism therapy, so technologies that focus on managing or tracking the information requirements associated with therapies should also be considered [18].

2.2 Triadic Interactions

Among the trends emerging from studies of using robots to interact with children with autism, one trend captures our attention because it seems most likely to allow robot-assisted therapies to generalize to other social domains. This trend is the emergence of triadic interactions between child, robot, and therapist. This triad is important because it suggests the possibility of using the robot as a tool in the hands of a skilled therapist to trigger social interactions between the child and the therapist. At least four researchers have noted the potential importance of this triadic interaction [6,9,14,24], and at least two have designed tools for a therapist that allows him or her to modulate robot behavior while interacting with children with autism [24,25].

We believe that interactions with the robot, though interesting, are less likely to produce generalizable social skill development than interactions between a child and the therapist. Indeed, robot social skills are very simple and are likely to remain simple for the near future, a fact that is perhaps the reason that robots trigger social responses in children with autism that are not usually triggered by human-human interactions. To support generalization, it is desirable (a) to use the robot as an engaging object of attention that allows therapist and child to engage in joint attention, (b) to shape social behaviors triggered by a robot by including a therapist "in the loop", or (c) to use the robot as a type of "cognitive orthotic" to facilitate social interactions between a child with autism and his or her caregiver, even if the child is unable to engage in social interactions in the absence of the robot.

We propose that effective therapies can be built on triadic relationships between social agents: robot, child, and another human. This other human can be a therapist, a teacher, a parent, or another child. There are two different triads that seem most promising, as illustrated in Figure 1. The first is between the child, the robot, and the child's Companion. This Companion can be a therapist, parent, teacher, or another child. This triad has been identified as potentially valuable in therapy and is known as robot-mediated interaction [11,12]. The second triad is between the robot, the child's Companion, and a "Wizard of Oz." Although this triad affects the robot's behavior, the presence of the Wizard is hidden from the child, as indicated by the wavy line between the robot and Wizard in Figure 1. The Wizard can extend the robot's perception and behaviors beyond what is possible using existing state-of-the-art algorithms, and can allow the robot to display a wider range of behaviors. The Wizard and the child's Companion can communicate via some means (radio, chat window, etc.) so that they can collectively work toward a therapeutic goal.

With the presence of the Companion and Wizard, it is possible to strongly emphasize robot-mediated interactions between a therapist and the child. The therapist can either play the role of Companion or Wizard, or perhaps both simultaneously [14]. We refer to the presence of a therapist modulating robot behavior as Therapist-in-the-Loop Assistive Robotics (TiLAR). We emphasize again that our objective is to use the robot as a facilitator in creating connections between the child and the therapist or other people, and the triadic therapies that we are exploring are designed to minimize the importance of the robot and emphasize the importance of other people in the triad.

2.3 Requirements for Interface Design

Given the constraints on the role of the robot in the interaction triad, it is desirable to understand the constraints on the user interface for interacting with the robot. Most prior work in this area has emphasized the interactions between robot and child, and there appears to be no prior work that systematically studies how to create a user interface to support the therapist. Since supporting the therapist by making the robot an effective and responsive tool in his or her hands is the goal of this work, we identify several practical constraints for this interface.

First, the user interface should allow interactions for either a Wizard or Companion location in the triad. For a Wizard, the user interface could allow a storyboard or other graphical user interface, but for a Companion the interface should probably be unobtrusive so that it does not distract the child. Indeed, controlling the robot unobtrusively was a key influence on the way the therapist managed robots in both Robins' and Scassellati's work [24,25].

Second, the user interface should be created for someone who is not a roboticist or computer professional. The therapist will be trained to understand and to help children, a task that often requires substantial mental workload on the part of the therapist. If the interface requires the therapist to do some form of "programming" during the therapy, it will likely be unusable.

Third, the user interface should support a specified clinical script, but be flexible enough to allow the therapist to modify behaviour in response to the child's responses. To track progress of a therapy and to measure effectiveness, robot-assisted therapies should have specific therapeutic goals and should exist within a framework specified by the training and schooling of the child. This influence tends to constrain robot behaviours to a fixed set that may be rigid and repeatable. However, the flexibility of having a semi-autonomous, socially responsive robot in a therapeutic loop means that unplanned behaviours may be exhibited by the child, and the therapist should have flexibility to respond to those unscripted interactions.

Fourth, the user interface should allow the set of useful robot behaviours to grow. This could include using interactive learning, wherein the robot's behaviours are modified over time by feedback from the child or the therapist. This could also

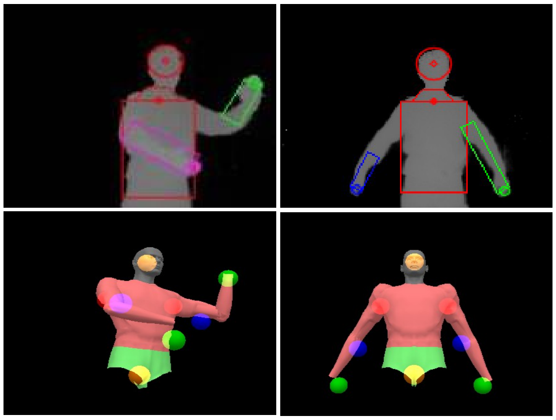


Figure 2. Avatar imitation (bottom) from estimated human pose (top).

include having the human demonstrate behaviours to the robot, which are then added to the robots behaviour repertoire. Finally, this could include the ability to sequence primitive behaviours into a more sophisticated social script. This could be especially beneficial for high functioning children, such as those with SLI.

We are developing a computer interface to provide to clinicians the functionality just described. Specifically, it will allow therapists to design and record robot behaviors that may be played back during therapy sessions, either as sequences of behaviors (storylines) or as stand-alone actions that the therapist can specify in real time. Additionally, the therapist will provide the means by which clinicians can initiate, terminate, and modulate the imitation behavior of the robot or avatar in automated imitation therapies, which will be discussed in the next section.

3 IMITATION THERAPY

Imitation therapy has been shown to provide potential benefits as a tool in the clinical treatment of autism [28]. One form of this therapy involves a therapist or other person imitating some or all of the behaviours of the child with autism. In this form (therapist imitating child), imitation therapy for autistic children has been expanded into the field of robotics, with the therapist directly controlling the robot's motion to imitate the motion of the child [14,17,29]. In a clinical setting, direct robot control would place a prohibitive workload on the therapist, thereby limiting his or her ability to interact with and observe the child.

The members of our research team at Honda Research Institute have developed algorithms and software that may enable real-time, automatic imitation of children with autism. The foundation for this work is the algorithm that enables real-time estimation of human pose from 3D time-of-flight data from a single depth camera sensor [30]. The method is markerless, and does not require an *a priori* model of the human whose pose is to be imitated. The system is capable of estimating pose of a 17 DOF upper body representation at a rate of 10 frames per second on a standard PC. The problem of transferring estimated human pose data to a humanoid robot has also been addressed [7,19] and demonstrated on avatars (Figure 2) and physical robots, including ASIMO (Figure 3).

The ability to estimate and imitate the motion of children automatically may be a valuable tool in autism treatment. To validate this tool, we propose to conduct studies with typical children and children with autism, in which we explore the set of behaviours that may be elicited through avatar- and robot-based imitation involving ASIMO and a custom-built robot whose design is described below.

In a research setting, our interface will allow clinicians to change the behavior of the imitating robot or avatar. For example, therapists will be able to cause the system to amplify or attenuate the imitated motions as a way of investigating the responses that robot imitation elicits in children. The therapist will also be able to disable certain degrees of freedom, change the imitating avatar, or introduce a delay in the imitated motions. This last feature was identified by our team of clinical researchers as an important component in an automated imitation suite; in their experience, children with autism require additional time to process imitated motions.

In addition to the obvious questions related to eliciting behaviors through robot-based imitation therapy, there are other open research questions that must be addressed for this type of system to be used in clinical settings. Calibration of the pose estimation system presents an interesting challenge for children with autism, who may not be able to remain immobile in the calibration pose for the few seconds required by the software. We are investigating the use of a game to help the child to assume the proper pose. Additional challenges include ensuring that the child remains within the usable volume of the camera, and segmenting out therapists or other participants that must pass into the field of view to assist the child.

4 ROBOTIC TOOLS TO ASSIST CLINICIANS

As mentioned previously, robots have the potential to induce behaviours in children with autism that are difficult to elicit in other ways. One important question related to robot-based therapy is how to create robots that elicit desired behaviors repeatably and that encourage generalization/transfer of those behaviors into everyday life. Our work seeks to explore this issue, in part, by developing a robot that (a) encourages joint attention between children with autism and clinicians, parents, etc., and (b) is easily adaptable to enable us to explore the role of robot appearance in the transfer of behaviors. Details of these objectives are discussed below.

4.1 Related Work

Considerable research has been published on the design of

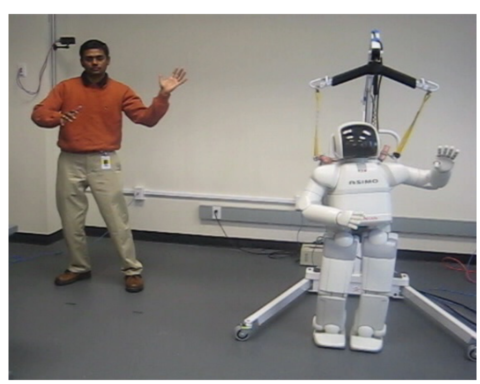


Figure 3. Real-time robot-based imitation using ASIMO.

expressive social robots, although the design and application of robots to autism therapy and diagnosis has been more limited. A brief overview of some of the related work will now be presented. Based on their studies using Robota, a small robot with the appearance of a toy doll [14], researchers at the University of Hertfordshire concluded that the best design for a robot to be used in interactions with autistic children would be one that is child-sized, has the general characteristics of a human face, but that is not so detailed as to overwhelm the children [17]. They have since created the robot KASPAR, which follows these criteria. Researchers at the National Institute of Information and Communications Technology in Japan developed the mechanical-looking robot Infanoid for use in human-robot interaction studies. In studies with children with autism, they observed that the children often focused more on the robot's mechanical parts than on the social interaction. The robot Keepon, with its simple, snowman-like design, was developed to address this concern [20]. Tito, a cartoon-like humanoid robot, was developed at the University of Sherbrook to combine the benefits of interacting with anthropomorphic forms with the appeal of less realistic robots [21]. To assist them in teaching complex social skills to high-functioning children with autism, researchers at the University of Pisa have developed the Face robot, which is designed to appear very human and realistically represent human emotions [22]. Other researchers have used non-humanoid, mobile robots to develop therapies for children with autism [13, 26]. For example, the toy dinosaur robot Pleo has been used to engage children with autism in triadic interactions with researchers [24].

4.2 Requirements for Robot Design

A robot designed to support the therapies proposed in this paper should be capable of general motions to enable imitation therapies and general behaviors choreographed by the therapist. Furthermore, based on discussions with our colleagues from the fields of psychology and communication disorders, it would be highly desirable to create a robot capable of engaging children with autism in exercises designed to encourage joint eye gaze and response to pointing [8]. Specifically, the robot should have the ability to point at an object or person with its arm and simultaneously direct its gaze to the same location. A therapist will be able to command the robot to point and gaze at a number of pre-programmed objects or locations in the room; the therapist will then attempt to engage the child in joint attention activities centered on the robot's pointing behaviors.

With these capabilities in mind, we are developing a small upper-body anthropomorphic robot with 4 degrees of freedom (DOFs) in each arm to enable general motions (including pointing), a 2-DOF neck to enable general gaze directions, and a simple cartoon-like mechanical face to represent basic emotions. We acknowledge that the capabilities of such a robot do not differ significantly from systems developed by other researchers. However, when combined with the pose estimation and imitation capabilities described previously, the system we are developing will provide unique capabilities as a research tool. Moreover, the system's ability to coordinate eye gaze and arm pointing will enable us to explore interesting questions related to joint attention and nonverbal social cues as they relate to children with autism.

Another facet of our research is to investigate if any advantage may be gained by replacing a typical mechanical robot face with a computer-generated avatar face displayed on a small monitor mounted to the robot neck. This approach has been demonstrated by researchers at Carnegie Mellon University with general-purpose robots such as GRACE [23], although we are not aware of this approach having been taken with robots for interacting with children with autism. We have chosen to include this feature in the robot under development because a computer-generated face will allow us to explore more thoroughly the relationship between appearance and clinical benefit, while still maintaining the physical embodiment that seems to be one source of the appeal of robots to children with autism. We acknowledge that moving away from a 3D face to a 2D screen may present challenges when interacting with children with autism; studies using this robot/face combination will allow us to ascertain the nature of those challenges. Furthermore, for children with SLI, interacting with a robot with a 2D face may not create the same type of challenges as it would for children with autism.

5 MOVING FORWARD

The use of robots as therapy tools for the treatment of children with autism and SLI represents an exciting and promising direction in the field of HRI. The research described in this paper is designed to move this work forward by developing new tools for therapists, in the form of novel interfaces, robot platforms, and algorithms for imitation therapy. The aim of these robot tools is to improve the effectiveness of treatment for children with autism and SLI by bringing to the forefront the relationship between child and therapist. Working with colleagues at Honda Research Institute and in the fields of psychology, communication disorders, computer science, and mechanical engineering, we will continue to develop these tools through clinical trials with children with autism and SLI at the Communication Disorders Clinic at Brigham Young University.

ACKNOWLEDGMENTS

This work is supported by a 2008 Honda Initiation Grant and a Mentored Environment Grant from Brigham Young University.

REFERENCES

[1] K.I. Hart, M. Fujiki, B. Brinton, and C.H. Hart. The Relationship Between Social Behavior and Severity of Language Impairment. *Journal of Speech, Language, and Hearing Research*, 47: 646-662 (2004).

[2] M. Fujiki, M.P. Spackman, B. Brinton, and A. Hall. The Relationship of Language and Emotion Regulation Skills to Reticence in Children with Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, 47: 637-646 (2004).

[3] M.P. Spackman, M. Fujiki, and B. Brinton. Understanding Emotions in Context: The Effects of Language Impairment on Children's Ability to Infer Emotional Reactions. *Int. Journal of Language and Comm. Disorders,* 41(2): 173-188 (2006).

[4] M. Fujiki, M.P. Spackman, B. Brinton, and T. Illig. The Ability of Children with Language Impairment to Understand Emotion Conveyed by Prosody in Narrative Passage. *Int. Journal of Language and Comm. Disorders,* 43(3): 330-345 (2008).

[5] B. Brinton, M.P. Spackman, M. Fujiki, and J. Ricks. What Should Chris Say? The Ability of Children with Specific Language Impairment to Recognize the Need to Dissemble Emotions in Social Situations. *Journal of Speech, Language, and Hearing Research,* 50: 798-811 (2007).

[6] H. Kozima, C. Nakagawa, and Y. Yasuda. Interactive robots for communication-care: a case-study in autism therapy. *Procs. IEEE Int. Symp. on Robot and Human Ineractive Communication* (RO-MAN05), Nashville,

USA, (2005).
[7] B. Dariush, M. Gienger, A. Arumbakkam, C. Goerick, Y. Zhu, and K. Fujimura. Online and markerless motion retargeting with kinematic constraints. *Procs. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems* (IROS 08), Nice, France (2008).
[8] B. Brinton, M. Fujiki, L. Robinson, and M. South. Personal communication.
[9] A. Duquette, F. Michaud, and H. Mercier. Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. *Autonomous Robots – Special Issue on Socially Assistive Robotics* 24(2):147-157 (2007).
[10] F. Tanaka, A. Cicourel, and J.R. Movellan. Socialization between Toddlers and Robots at an Early Child Education Center. *Procs. of the National Academy of Sciences of the United States of America*, 104:17954-8 (2007).
[11] C.M. Stanton, P.H. Kahn Jr., R.L. Severson, J.H. Ruckert, and B.T. Gill. Robotic Animals Might Aid in the Social Development of Children with Autism. *Procs. ACM/IEEE Int. Conf. on Human-Robot Interaction*, Amsterdam, Netherlands (2008).
[12] B. Robins, P. Dickerson, P. Stribling, and K. Dautenhahn. Robot-Mediated Joint Attention in Children with Autism: A Case Study in Robot-Human Interaction. *Interaction Studies*, 5(2): 161-198 (2004).
[13] D.J. Feil-Seifer and M. Mataric. Robot-Assisted Therapy for Children with Autism Spectrum Disorders. *Procs. Conf. on Interaction Design for Children: Children with Special Needs*, Chicago, USA (2008).
[14] B. Robins and K. Dautenhahn. The Role of the Experimenter in HRI Research—a Case Study Evaluation of Children with Autism Interacting with a Robotic Toy. *Procs. IEEE Int. Symp. on Robot and Human Ineractive Communication* (RO-MAN06), Hatfield, UK (2006).
[15] B. Scassellati. How Social Robots Will Help Us to Diagnose, Treat, and Understand Autism. *Procs. 12th Int. Symp. on Robotics Research*, San Francisco, USA (2005).
[16] B. Robins, K. Dautenhahn, and J. Dubowsky. Robots as Isolators or Mediators for Children with Autism? A Cautionary Tale. *Procs. Social Intelligence and Interaction in Animals, Robots and Agents* (AISB05), Hatfield, UK (2005).
[17] B. Robins, K. Dautenhahn, and J. Dubowsky. Does Appearance Matter in the Interaction of Children with Autism with a Humanoid Robot? *Interaction Studies*, 7(3): 479-512 (2006).
[18] J.A Kientz, S. Boring, G.D. Abowd, and G.R. Hayes. Abaris: Evaluating Automated Capture Applied to Structured Autism Interventions. *Procs. Int. Symp. on Ubiquitous Computing.* Tokyo, Japan (2005).
[19] B. Dariush, M. Gienger, B. Jian, C. Goerick, and K. Fujimura. Whole body humanoid control from human motion descriptors. *Procs. IEEE Int. Conf. on Robotics and Automation* (ICRA 08), Pasadena, USA (2008).
[20] H. Kozima, C. Nakagawa, and Y. Yasuda. Interactive robots for communication-care: a case-study in autism therapy. *Procs. IEEE Int. Workshop on Robotics and Human Interactive Communication* (2005).
[21] A. Duquette, H. Mercier, and F. Michaud. Investigating the use of a mobile robotic toy as an imitation agent for children with autism. [Online] http://www.csl.sony.fr/epirob2006/posters/Duquetteetal.pdf.
[22] G. Pioggia, R. Igliozzi, M. Ferro, A. Ahluwalia, F. Muratori, and D. De Rossi. An android for enhancing social skills and emotion recognition in people with autism. *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, 13(4): 507-15 (2005).
[23] R. Simmons, D. Goldberg, A. Goode, M. Montemerlo, N. Roy, B. Sellner, C. Urmson, M. Bugajska, M. Coblenz, M. Macmahon, D. Perzanowski, I. Horswill, R. Zubek, D. Kortenkamp, B. Wolfe and B. Maxwell. GRACE: An Autonomous Robot for the AAAI Robot Challenge. *AI Magazine* (2003).
[24] B. Scassellati. Personal communication.
[25] B. Robins. Personal communication.
[26] K. Dautenhahn and I. Werry. Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & Cognition*, 12:1-35 (2004).
[27] G. Conti-Ramsden, Z. Simkin, and N. Botting. The prevalence of autistic spectrum disorders in adolescents with a history of specific language impairment. *Journal of Child Psychology and Psychiatry*, 47:621-628 (2006).
[28] T. Field, C. Sanders, and J. Nadel. Children with autism display more social behaviors after repeated imitation sessions. *Autism*, 5:317-323 (2001).
[29] H. Kozima and C. Nakagawa. Interactive robots as facilitators of children's social development. *Mobile Robots Towards New Applications*, A. Lazinica, ed., pp. 784 (2005).
[30] Y. Zhu, B. Dariush, and K. Fujimura. Controlled human pose estimation from depth image streams. *CVPR Workshop on Time of Flight Camera Based Computer Vision*, Anchorage, USA (2008).

Proceedings New Frontiers in Human-Robot Interaction, K. Dautenhahn (Ed.), symposium at the AISB09 convention, 8-9 April, Edinburgh, Scotland

Creating Trustworthy Robots: Lessons and Inspirations from Automated Systems

Munjal Desai[1], **Kristen Stubbs**[1], **Aaron Steinfeld**[2] and **Holly Yanco**[1]

Abstract. One of the most significant challenges of human-robot interaction research is designing systems which foster an appropriate level of trust in their users: in order to use a robot effectively and safely, a user must place neither too little nor too much trust in the system. In order to better understand the factors which influence trust in a robot, we present a survey of prior work on trust in automated systems. We also discuss issues specific to robotics which pose challenges not addressed in the automation literature, particularly related to reliability, capability, and adjustable autonomy. We conclude with the results of a preliminary web-based questionnaire which illustrate some of the biases which autonomous robots may need to overcome in order to promote trust in users.

1 Introduction

Effective human-robot interaction depends not just on the design of the interaction but also on the level of trust that the user has in the robotic system. In their survey of trust and automation research, Lee and See define trust as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" [20]. The level of trust that people have in an automated system is a key factor that influences their use of that system. An inappropriate level of trust may result in inappropriate use (misuse) or disuse of automation, which may result in poor performance [20]. While considerable work has been done on trust in automation, we feel that the added uncertainty and vulnerability inherently present in robots necessitates dedicated work on trust and robotics.

People who place little trust in automation may disuse it. This lack of usage has potential safety implications, as the automation may be better equipped to handle crises than they are. For example, a crash avoidance system in a car may be disabled or tampered with if the driver considers past behaviors untrustworthy, yet the crash avoidance system might be able to react more quickly than the driver in some situations (e.g., [41]). As we have observed during the course of our previous studies in human-robot interaction (HRI), robot users who do not trust a robot often disengage its autonomous capabilities, such as obstacle avoidance. This avoidance of automation, particularly when reverting to direct control (teleoperation), may lead to damage of both the robot and its environment. For example, Yanco et al. observed and evaluated robotic systems designed for urban search and rescue (USAR) at the Robot Rescue Competition at AAAI 2002 [45]. During one of the system's runs, there was a clear Plexiglas sheet present in the path of the robot. The sensors on the robot detected the Plexiglas, and because the robot was utilizing automation, the robot would not drive forward even when the user tried to force it to drive forward. This frustrated the user, who switched to manual control and drove through the Plexiglas sheet.

Problems can also arise when people place too much trust in an automated system. For example, robot users may incorrectly internalize a certain level of trust (e.g., that the robot will not collide with nearby humans) and inadvertently place themselves or others in harm's way. For example, during the experiments carried out by Desai [6], there was an expert user who had been trained with the system and was aware of the system's capabilities and shortcomings. This user decided to use the maximum level of autonomy supported by the system. The robot performed poorly, but the user kept the system under full automation for more than half the run. One of the statements that the user made before starting the run indicated that the user was very enthusiastic about autonomy and wanted to try it out. This statement indicated a very high level of initial trust in the system. This attitude combined with the fact that the user received little feedback about the performance of the system resulted in a situation in which the user's trust was largely misplaced and the robot made numerous collisions. One of the factors that made judging the robot's performance difficult from the user's perspective was the fact the user and the robot were not collocated, so the user could not easily observe the robot or its environment. This separation is not normally the case with most automation systems.

Advances in social robotics may exacerbate situations in which users mistrust a system. Related work by Parasuraman and Miller [33] described how automation etiquette has an impact on user trust and overall human-automation performance in traditional automation tasks. Their study showed that "good automation etiquette can compensate for low automation reliability," suggesting that people place more trust in a polite robot than is warranted by the robot's actual abilities. In addition, van Mulken [44] found that displaying information in a personified manner did not affect trust. This indicates that designers of social robots may need to consider how their robot presents information (what level of politeness and personification) in order to foster appropriate levels of trust in the system.

Due to potential problems with trusting a robotic system too much or too little, it is important to develop a model that will allow a robotic system to estimate its user's current level of trust. In this paper, we present an overview of previous work in trust in automated systems and discuss specific areas in which robotics poses challenges not previously addressed by this body of work. We also discuss a preliminary web-based questionnaire which we conducted to examine people's attitudes toward robotic automation.

[1] University of Massachusetts Lowell, One University Avenue, Lowell, MA 01854. Email: {mdesai,kstubbs,holly}@cs.uml.edu
[2] Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. Email: steinfeld@cmu.edu

2 Research on Trust in Automation

Parasuraman and Riley define automation as "the execution by a machine agent (usually a computer) of a function that was previously carried out by a human" [34]. Automation has traditionally been employed in systems that are complicated, tedious, or time critical, but it has also been used for economic reasons [34]. When automation was first introduced in the 1930's, its use was limited to large industries; however, at the present, automation can be found in many places, from home appliances to the Mars rovers.

Automation has always had weaknesses: namely, it has only been effective in well-structured and controlled environments and continues to remain so. To avoid catastrophic failures in safety critical systems due to either flaws or limitations of automation, an operator must be present at all times to take control of the system. Situations of this kind in which a human operator is working with an automated system are referred to as "human-in-the-loop control." While utilizing a human operator may be beneficial in certain situations, addressing the inadequacies of automation for the human-in-the-loop control creates a different set of problems. When an operator is added to the system, improving overall system performance requires more than simply optimizing operator performance and, separately, optimizing automation performance. The interaction between the two needs to be considered as well.

For several decades, researchers in the automation field have examined the factors which influence people's trust of automated systems and how this level of trust, in turn, effects the way in which people use, misuse, or disuse automation. Researchers have shown that trust influences operators' use of automation (e.g., [8, 19, 31]): the more operators trust automation, the more they tend to use it. Moreover, if an operator trusts his own abilities more than those of the automated system, he tends to choose manual control; however, if an operator trusts the automation more than his own capabilities, he is more likely to choose automation over manual control.

Not only has it been demonstrated that a user's level of trust affects how much he will rely on an automated system, but numerous studies have also been conducted to examine the factors which influence this level of trust (see [20] for an overview). Specifically, Dzindolet et al. [8] demonstrated the impact of system performance on user trust. The results of this study indicated that while users initially placed trust in a decision aid and agreed with its suggestions, as users observed the system making errors they would distrust even a generally high-performing aid unless provided with reasons as to why the errors had occurred. However, providing this type of information increased trust in the automated aid even when the aid performed poorly. Besides the performance of the system, additional factors contributing to a user's trust of an automated system include the recency of errors made by the system, the user's prior knowledge about the system's performance, the user's knowledge of the capabilities of the system, and the user's expectations of the system's performance [39].

Different models have been hypothesized regarding how these different factors influence each other and ultimately the operator's reliance on automation. One such model was proposed by Riley [37] and is shown in Figure 2. The dashed lines indicate the unproven hypotheses and the solid lines indicate relationships that have been proved by experiments. This model, like most, does not consider some factors that are relevant to robots such as interface usability, proximity to robot (co-located or remote-located), situation awareness, dynamics of operating environment, etc. These factors heavily influence automation in robotics. Since the performance of automation has an effect on users' trust of automation it is important to consider these factors. Of these factors, some work has been done on interface usability [2] and situation awareness [23].

Figure 1 shows a generic trust model which has been augmented with factors more relevant to robotics. Many researchers have proven that the way the user allocates control or uses automation depends on the amount of trust that the user has in the automation and the amount of trust that user has in his own capabilities (e.g., [5, 8, 19, 21, 37]). A user's trust of his own capabilities is most often referred to as "self-confidence." Some studies have reported a certain amount of lag between changes in trust and self-confidence and an actual change in allocation strategy; this lag is referred to as "inertia" [21]. Atoyan et al. found that interface design plays an important role in influencing users' trust in automation [2].

When the user changes the allocation strategy, the performance of the system inevitably changes. For the feedback loop to close, the user needs to observe this change in performance. Depending on the system, there might be a significant time delay before the user observes the change in performance. This time delay may result because only cumulative feedback is provided to the user [8], or it may be a result of the time required to send information to the user over the communication channel. The significance of a particular amount of time delay depends on the nature of the system. For example, a delay of a few hundred milliseconds will have a drastic effect on a USAR system but may have little effect on a Mars rover. The change in system performance is also dependent on machine accuracy, and, as explained in Section 3.1, the performance of automation in robotic systems is generally not very high.

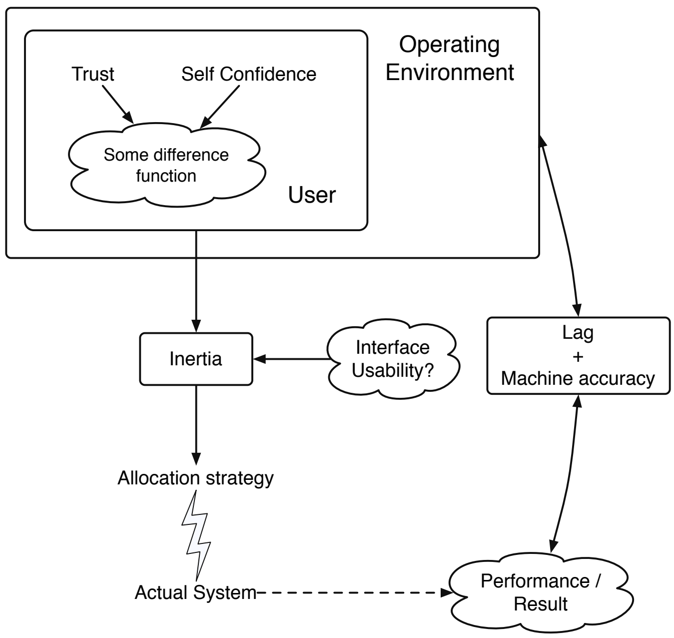


Figure 1. A basic model of trust adapted for robotics. It shows that trust and self-confidence influence automation allocation along with some factors that would be more relevant to robotics like interface usability, lag, and machine accuracy.

Table 3 at the end of this paper lists previous work related to trust and automation. For a more comprehensive coverage of trust and au-

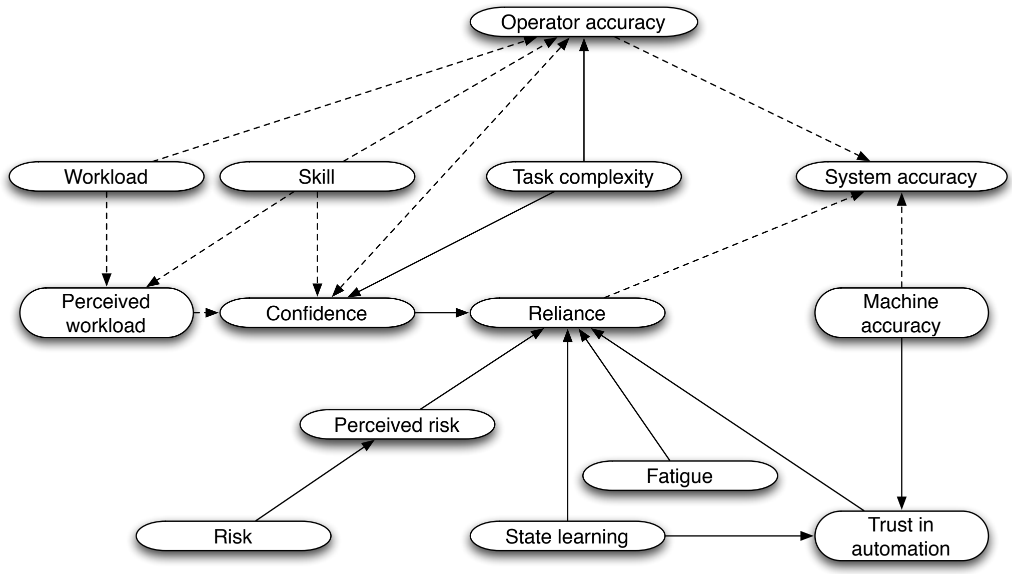


Figure 2. Factors influencing automation use, according to Riley [37]. The dashed lines indicate the relationships that are yet to be proven and the solid lines indicate the relationships that have been experimentally proven.

tomation, see Lee and See [20]. Table 3 lists experimental setup details such as the number of participants, background of participants, and experiment task. It also lists the main contribution of the papers. Most of the experimental setups were some form of automated decision aids. There were two studies that were based on questionnaires ([16, 26]). Most of the experiments were modeled after systems that were not very complex, such as orange juice pasteurization [19], character identification [37], camouflaged soldier detection [9], etc. Most of the studies recruited students. One study in which pilots participated was [37]. It is important to note that while most of the studies utilized simulations, some of the simulations were modeled after real systems ([19, 28, 30]) while others were not.

In one of their experiments with simulated pasteurization plants, Lee and Moray [22] found that automation allocation was dependent on trust and self-confidence. They also found that operators exhibited inertia, that there was a bias towards manual control, and that this bias was even more prominent during the initial stages. In another experiment, Lee and Moray also found that automation usage is dependent on individual biases [19]. The bias towards manual control was also validated by Riley in his experiments with a character identification system [37]; due to these findings in a small sample set of users, he suggests running larger sets of participants.

As can be seen in Table 3, most trust in automation studies have had relatively few participants. Riley also conducted similar experiments with students and pilots and found that automation allocation strategy followed the same pattern for both groups except that the pilots relied on automation more than the students did. This difference raises the question of testing systems with domain experts, which is seldom done in automation research but is standard in human-computer interaction for usability studies.

Dzindolet et al. conducted experiments with an automated decision aid for camouflaged soldier detection [9]. In one of their experiments, they found that participants who had little information about the reliability of the system considered the system trustworthy. This bias is referred to as the “positivity bias”. The concept of the positivity bias was developed in the field of social psychology; the idea is that people tend to be biased to think highly of other people when adequate information about them is not available. If participants had a positivity bias towards an automated system, the participants would be more likely to use automated control (at least initially). However, Dzindolet’s findings contradict those of Lee and Moray [22] and Riley [37], in which users showed an initial bias towards manual control.

3 Automation and Trust in Robotics

While the broad range of automation research provides a context for examining issues of trust with robots, there are a number of factors that limit how well this work generalizes to the robotics domain. For example, studies of automated systems have tended to utilize systems such as autopilots, flight management systems, vision systems for target or obstacle identification, and factory control systems [20]. Participants interact with a simulated system, which allows experimenters to inject errors and observe how participants’ level of trust of and use of the system change as a result. The systems used in these experiments generally have no physical embodiment and do not interact with the physical world. Furthermore, these automated systems tend to be designed for rigid tasks; that is, each system performs only one very specific type of task. Robots are generally designed to accomplish a wider range of tasks: an assistive robot might be needed to fetch a cup of coffee from the kitchen one day and a hat from the closet the next, and a robotic system for USAR can be deployed at different disaster areas with different physical layouts and characteristics, such as a collapsed building or a mine shaft. In fact, it is quite common for robots to be used for tasks their designers never envisioned. Thus, the implications of these automation studies for physically embodied robots with noisy sensors operating in dynamic environments are less clear.

3.1 Automation Reliability

Reliability is one characteristic which has been shown empirically to affect a user's trust in a system [5, 9, 37]. Designing automation for robotics is more challenging than for other automated systems because of the difficulties in modeling the robot and its environment and the challenges posed by poor sensor data. Designing automation requires modeling an existing system. Regardless of the implementation architecture, building such a model requires that all possible states be mapped to an action. This mapping is easy to do for traditional automated systems, such as the orange juice pasteurization plant utilized in Lee and Moray's experiments [19]; however, creating this mapping is difficult to do when the number of possible states is very large or dynamic. Designing automation for such systems requires many approximations, reducing the reliability of automation even in the presence of perfect sensor information. Robots designed for urban search and rescue, assistive technology, or unmanned surveillance must operate in unknown, unstructured, and dynamic environments. This lack of environmental constraints makes designing automation to cover all possible circumstances the robot might encounter very difficult. As a result, these types of robotic systems are likely to have lower reliability than other automated systems.

Automation performance not only depends on the implementation of automation by the designers, but it is also heavily dependent on information from sensors. Most automation experiments conducted regarding trust have been in simulation (e.g., Table 3). In these experiments, the reliability of automation is artificially controlled; most often, the reliability is constant throughout the entire experiment or constant for relatively large time periods. This type of experimental design may help to highlight the effect that reliability has on trust and control allocation; however, since the resulting models are derived from simulated systems, their applicability to real, physical systems remains unclear. This issue is very important in robotics because most sensor modalities used are either unreliable (i.e., sonar, infrared, etc.) or their accuracy is dependent on environmental factors (such as light, reflectivity of surfaces, other characteristics of surfaces, etc). This suggests that at the very least, for robotics, simulated experimental setups with reliability modeled after real world systems or real world experiments are needed. While studies have been conducted which examine the effects of reliability on trust, given the dynamic nature of reliability in robotics, a comprehensive study of the effects of reliability in robotics is required to either validate existing trust models or modify them.

3.2 Automation Capability

Automation capability defines to what extent system operations can be controlled by automation, whereas automation reliability defines how well system operations can be performed. In most previous automation experiments, automation capability has not been an issue; however, in robotic systems, the maximum capability of automation must be considered separately from reliability.

In most automation experiments conducted to date, the automated system is capable of performing the entire operation. This capability is due in part to the oversimplified nature of the experimental setup. An example would be controlling the valves in a pasteurization plant or correctly indicating the presence of a camouflaged soldier in an image. From the operator's perspective, the mental model of the automation's capability is relatively static: the automation always has the capability to complete the task at hand; however, this may not be the case with robotics. In several robotic domains, systems employ discrete levels of autonomy for different reasons. These levels of autonomy define what tasks the automation can perform. Parasuraman et al. have described a classic example illustrating how different levels of autonomy can be utilized [35]. USAR systems developed by INL [3], and UML [17] have several discrete autonomy levels, requiring the user to accurately learn, remember, and use the correct mental model for automation. The need for a complicated model of automation capability has the potential to drastically increase the chance of misplaced trust in automation. UML has also implemented sliding scale autonomy for USAR [6]. According to Desai and Yanco, a sliding scale autonomy system is a continuum of autonomy levels from full teleoperation to full autonomy [7]. From an automation point of view, a sliding scale system might be better than a discrete autonomy system, but the influence of such sliding scale autonomy on trust has yet to be studied.

3.3 Changing Levels of Autonomy

Furthermore, the fact that robotic systems may operate under changing levels of autonomy is generally not addressed in this literature. In some robotic systems, the user specifies the desired level of autonomy; in others, the robot may change its own level of autonomy without specific human direction to do so. Desai [6] lists factors in the robotics application domain that could govern the minimum as well as maximum amount of automation a system can have. Some robot application domains, such as urban search and rescue (USAR), can have very unstructured environments [29], which require the presence of a human operator to assist the automated system. A robotic system with different levels of autonomy requires that the user develop an allocation control strategy to decide how much autonomy the system should have at any given time. However, most of the research regarding allocation control strategy that has been done has been mainly in the field of industrial automation or aviation automation. Most often in such situations, the automation can either be turned on or turned off (e.g., [5, 21, 37]). Most autonomous robotic systems employ a discrete autonomy system in which there are several autonomy levels to choose from, which complicates the problem of automation allocation: the user must now decide not only whether or not automation must be used but also how much; thus the level of autonomy must also be considered as a factor related to the user's trust of the system.

In robotics, adjustable autonomy systems are sometimes referred to as mixed initiative autonomy (MIA) systems. However, the term "adjustable autonomy" is also used to refer to systems in which the operator and the automated system can both change the level of autonomy. A system like this was implemented by Desai [6]. In domains like USAR, in which the operator may utilize a robot for long periods of time, it is very difficult for the operator to maintain situation awareness the entire time. The widely accepted definition of situation awareness is "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" [10]. Situation awareness can be easily lost if the robot is operating autonomously. This lack of good SA might result in an accident if the automation fails for some reason. MIA systems can prevent these types of failures by gradually handing over control to the operator as the automation starts to fail. By transferring control to the operator gradually, the operator gains time to regain SA. The MIA system implemented by Desai [6] can also take over some amount of control from the operator if the automation detects that the operator is performing errors that the automation could avoid if it were in control. The effects of automation failures on trust in MIA systems have not been studied. Some researchers have tried to treat robots as peers rather than just tools (e.g., Marble et al. [25] and Fong et al. [11]). Such systems would implement some sort of MIA, which makes studying the effects of MIA on trust even more important.

3.4 Prior Work on Trust in Robotic Systems

To date, there has been little work examining issues of trust directly with robots, although some work has been conducted involving simulated robots. Dassonville et al. [4] conducted a study in which participants used a joystick to control a simulated PUMA arm. Errors were introduced into the simulation, and participants were asked to rate the reliability, performance, and predictability of the joystick's behavior (as well as how difficult it was to make such ratings). The results of the study were consistent with prior work in autonomous systems that suggest that the user's self-confidence is a significant factor which influences use of such systems.

More recently, Freedy et al. [13] have examined trust in the context of mixed-initiative command and control systems using the MITPAS (Mixed-Initiative Team Performance Assessment System) Simulation Environment. The researchers constructed a quantitative measure of trust by assuming that people use a rational decision model such that "trust behavior is reflected by the expected value of the decisions whether to allocate control to the robots on the basis of past robot behavior and the risk associated with autonomous robot control" [13]. For their evaluation, participants were asked to assume the role of a controller of an Unmanned Ground Vehicle (UGV); the UGV was able to autonomously target and fire, but participants were instructed to take control of the UGV if its autonomous behaviors would lead to a time delay or a failure. The experimenters varied the competency of the UGV's firing behavior and recorded participants' choices to override the UGV. The results suggested that participants who were able to ascertain whether the UGV was very competent or incompetent adjusted their behavior accordingly, seeming to trust the system to continue to maintain the same level of competence. In the case in which the system was of indeterminate competence, it was more difficult for users to adjust their behavior. Because the entire experimental setup took place in a video game-like environment, it is unclear how these results would generalize to physical robots.

While relatively little work has been done investigating trust in robots, there is a large body of research on trust in different types of technologies. Because we are interested in developing a model of trust for human-robot interaction, we have examined trust models that were developed for other technology domains. For example, Song et al. [40] developed a neural network-based trust model for understanding users' acceptance of recommendations from a system of heterogeneous agents. Another agent-related trust model was developed by Rehak et al. [36], who used fuzzy numbers to represent trust in cooperating ubiquitous devices. McKnight et al. [27] developed a trust model to understand users' acceptance of a website offering legal advice. However, all of these systems differ from robots in the same ways that the systems previously studied in the automation literature do. To advance the field of human-robot interaction, a systematic study of trust in human-robot interaction is necessary to build trust models in this domain.

4 Web-Based Questionnaire on Attitudes Towards Robotic Automation

To examine people's perceived level of comfort with robotic automation, we have conducted a preliminary web-based questionnaire. Participants were recruited through Mechanical Turk, a website which allows individuals and companies to post human intelligence tasks which are accepted and completed by online workers [42]. Participants were asked about a (fictitious) new car that had the ability to park itself automatically. In order to examine participant's initial biases towards the system, participants were not given any information about the competence or reliability of the fictitious car. Participants were asked to envision parking at a grocery store and to rank the following situations in order of how comfortable they would feel in each situation from 1 (most comfortable situation) to 6 (least comfortable situation):

- You park your car manually.
- Another driver manually parks their car next to your car.
- Another car automatically parks itself next to your car.
- Your car automatically parks itself (and you cannot override it).
- Your car automatically parks itself (but you can override it).
- You take a taxi and the taxi driver parks the taxi.

We received 176 responses to the questionnaire (69.3% female, 30.1% male, 0.6% unknown). Participants reported their ages as follows: 18 to 25, 22.1%; 26 to 35, 36.3%; 36 to 45, 22.1%, 46 or older, 18.1%; unknown, 1.1%. 97.7% of respondents reported having prior experience driving a car. The mean rankings of each scenario are shown in Table 1. We conducted a Friedman two-way analysis of variance to compare the rankings for the scenarios, which provided evidence for significant differences among the six rankings, $\chi^2(5) = 319.79$, $p < 0.001$. In order to determine which scenarios' rankings were significantly different from each other, we used Wilcoxon matched-pairs signed-ranks tests with the Bonferroni correction (Table 2).

Overall, 65% of respondents indicated they would be most comfortable manually parking their own car (mode rank = 1), and 55% of respondents indicated they would be least comfortable if the car was parking itself and they had no means to override it (mode rank = 6). While a taxi passenger also has no means to "override" a taxi driver, participants tended to feel less comfortable with the automated system. Similarly, 38.6% of participants ranked the situation in which another driver manually parks a car next to theirs as the second most comfortable (mode rank = 2), yet 36.4% of participants ranked the situation in which another car automatically parks itself next to theirs as their second to least comfortable (mode rank = 5), even though they would have no control over the other car in either case. This suggests that people may tend to trust a robotic system less than another person even in circumstances for which there may not logically be any difference in terms of the person's control over the situation.

Nomura et al. have demonstrated empirically that people's negative attitudes towards robots will affect their interactions with robots [32]. The types of biases which we observed in this survey represent negative attitudes which may impact how robotic automation is utilized. In order to improve HRI given these biases, a robotic system will need to adjust its interactions to align with the amount of trust placed in it by the user. For example, the robotic system could alter of warning thresholds (e.g., collision warnings or status warnings such as battery level), how often it asks for help, and which level of autonomy to use. Strategies for modeling human behavior within a robot have been examined (e.g., [12]), but more work is needed.

Table 2. p-values from Wilcoxon matched-pairs signed-ranks tests between the rankings for each scenario. All values less than 0.003 (adjusted from 0.05) indicate statistical significance.

Scenario	Self Manual	Another Driver Manual	Taxi	Self Auto : Override	Another Driver Auto	Self Auto : No Override
Self : Manual		$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
Another Driver : Manual			0.81	0.60	$p < 0.001$	$p < 0.001$
Taxi				0.48	$p < 0.001$	$p < 0.001$
Self : Auto : Override					$p < 0.001$	$p < 0.001$
Another Driver : Auto						$p < 0.001$

Table 1. Mean ranking, mode ranking, and percentage of participants whose responses matched the mode ranking, where 1 = most comfortable situation and 6 = least comfortable situation.

Scenario	Mean Rank	Mode Rank	Participants at Mode
Self : Manual	1.74	1	65.3%
Another Driver : Manual	3.31	2	38.6%
Taxi	3.36	3	27.3%
Self : Auto : Override	3.19	4	27.3%
Another Driver : Auto	4.36	5	36.4%
Self : Auto : No Override	5.04	6	55.1%

5 Conclusion and Future Directions

Prior work on trust in automated systems provides a foundation for understanding and modeling trust in human-robot interaction, but much work remains to be done. Because of the challenges of modeling a robot's sensors, actuators, and its environment as well as the challenges of interpreting noisy sensor data, automation for robotic systems is likely to be less reliable than the systems used in previous work on automation. In addition, a robot may be used in a variety of situations for a variety of tasks, and a robot may not always have the capability to complete every aspect of the task at hand. Thus, robotic systems may have a lower level of automation capability than the systems utilized in automation research. Adjustable autonomy and mixed-initiative robotic systems, including systems which may change their autonomy level dynamically, introduce additional complexity which may affect the user's trust in the system. Our web-based questionnaire also illustrates that users may be biased against robotic autonomy, even compared with situations in which there may be little difference in terms of their own control (i.e., a person riding in a car being parked by another person as opposed to by an automated system).

Further research is needed in order to create models of trust which are specifically tailored towards human-robot systems. The field of HRI should begin to investigate the question of trust through empirical studies, particularly relating to those factors which distinguish robots from other automated systems. Studies in which participants must execute a task with a real robotic system could include measures of perceived reliability and the system's actual reliability to compare how these factors influence participants' use of the system and reported trust of the system. For tasks in which robotic automation is only sometimes helpful, a careful examination of how participants understand the system's capabilities, and how this impacted trust in the system, would also be helpful. Examining the effects of changing levels of autonomy on trust, as well as the effect of automation failures, is another possible research area. In order to build systems which promote appropriate levels of trust, HRI designers will need to consider how to design both the robot's form and its interactions such that it provides feedback which will help the user understand the robot's capabilities and limitations.

Acknowledgments

This work is supported by the National Science Foundation (IIS-0546309).

REFERENCES

[1] S. Antifakos, N. Kern, B. Schiele, and A. Schwaninger, 'Towards improving trust in context-aware systems by displaying system confidence', in *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services*, pp. 9–14. ACM New York, NY, USA, (2005).

[2] H. Atoyan, J.R. Duquet, and J.M. Robert, 'Trust in new decision aid systems', in *Proceedings of the 18th International Conferenceof the Association Francophone d'Interaction Homme-Machine*, pp. 115–122. ACM New York, NY, USA, (2006).

[3] D.J. Bruemmer, D.D. Dudenhoeffer, and J. Marble, 'Dynamic autonomy for urban search and rescue', in *2002 AAAI Mobile Robot Workshop, Edmonton, Canada, August*, (2002).

[4] I. Dassonville, D. Jolly, and A. M. Desodt, 'Trust between man and machine in a teleoperation system', *Reliability Engineering and System Safety*, **53**, 319–325, (1996).

[5] P. de Vries, C. Midden, and D. Bouwhuis, 'The effects of errors on system trust, self-confidence, and the allocation of control in route planning', *International Journal of Human-Computer Studies*, **58**(6), 719–735, (2003).

[6] M. Desai, *Sliding scale autonomy and trust in human-robot interaction*, Master's thesis, University of Massachusetts Lowell, May 2007.

[7] M. Desai and H.A. Yanco, 'Blending human and robot inputs for sliding scale autonomy', in *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pp. 537–542, (2005).

[8] M. Dzindolet, S. Peterson, R. Pomranky, L. Pierce, and H. Beck, 'The role of trust in automation reliance', *International Journal of Human-Computer Studies*, (2003).

[9] M.T. Dzindolet, S.A. Peterson, R.A. Pomranky, L.G. Pierce, and H.P. Beck, 'The role of trust in automation reliance', *International Journal of Human-Computer Studies*, **58**(6), 697–718, (2003).

[10] M. Endsley, 'Design and evaluation for situation awareness enhancement', in *Human Factors Society, Annual Meeting, 32 nd, Anaheim, CA*, pp. 97–101, (1988).

[11] T. Fong, I. Nourbakhsh, C. Kunz, L. Fluckiger, J. Schreiner, R. Ambrose, R. Burridge, R. Simmons, L.M. Hiatt, A. Schultz, et al., 'The peer-to-peer human-robot interaction project', *Space*, **6750**, (2005).

[12] T. Fong, C. Thorpe, and C. Baur, 'Collaboration, dialogue, and human-robot interaction', in *Proceedings of the 10th International Symposium of Robotics Research*, London. Springer-Verlag.

[13] A. Freedy, E. DeVisser, G. Weltman, and N. Coeyman, 'Measurement of trust in human-robot collaboration', in *Proceedings of the 2007 International Conference on Collaborative Technologies and Systems*, (2007).

[14] A. Glass, D.L. McGuinness, and M. Wolverton, 'Toward establishing trust in adaptive agents', in *Proceedings of the 13th international conference on Intelligent user interfaces*, pp. 227–236. ACM New York, NY, USA, (2008).

[15] J.L. Herlocker, J.A. Konstan, and J. Riedl, 'Explaining Collaborative Filtering Recommendations', (2000).

[16] J.Y. Jian, A.M. Bisantz, C.G. Drury, and J. Llinas, 'Foundations for an Empirically Determined Scale of Trust in Automated Systems', (1998).

[17] Brenden Keyes, *MS Thesis: Evolution of a Telepresence Robot Interface*, Master's thesis, University of Massachusetts Lowell, 2007.

[18] M.T. Khasawneh, S.R. Bowling, X. Jiang, A.K. Gramopadhye, and B.J. Melloy, 'A Model for Predicting Human Trust in Automated Systems', *Origins*, (2003).

[19] J. Lee and N. Moray, 'Trust, self-confidence and supervisory control in a process control simulation', in *Proceedings of the Conference on Systems, Man, and Cybernetics*, (1991).

[20] J. Lee and K. See, 'Trust in automation: designing for appropriate reliance.', *Human Factors*, **46**, 50–80, (2004).

[21] J. D. Lee and N. Moray, 'Trust, self-confidence, and operators' adaptation to automation', *International Journal of Human-Computer Studies*, **40**(1), 153–184, (1994).

[22] J.D. Lee and N. Moray, 'Trust, self-confidence, and operators' adaptation to automation', *International Journal of Human-Computer Studies*, **40**(1), 153–184, (1994).

[23] C.L. Liu and S.L. Hwang, 'Evaluating the Effects of Situation Awareness and Trust With Robust Design in Automation', *International Journal of Cognitive Ergonomics*, **4**(2), 125–144, (2000).

[24] P. Madhavan, D.A. Wiegmann, and F.C. Lacson, 'Automation Failures on Tasks Easily Performed by Operators Undermine Trust in Automated Aids', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, **48**(2), 241–256, (2006).

[25] J.L. Marble, D.J. Bruemmer, D.A. Few, and D.D. Dudenhoeffer, 'Evaluation of supervisory vs. peer-peer interaction with human-robot teams', in *Proceedings of the 37th Annual Hawaii International Conference on Systems Sciences*, (2004).

[26] J. Masthoff, 'Computationally Modelling Trust: An Exploration', in *Proceedings of the SociUM workshop associated with the User Modeling conference, Corfu, Greece*, (2007).

[27] D. H. McKnight, V. Choudhury, and C. Kacmar, 'The impact of initial consumer trust on intentions to transact with a web site: a trust building model', *Journal of Strategic Information Systems*, **11**(3–4), 297–323, (2002).

[28] S.M. Merritt and D.R. Ilgen, 'Not All Trust Is Created Equal: Dispositional and History-Based Trust in Human-Automation Interactions', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, **50**(2), 194–210, (2008).

[29] Mark Micire, 'Evolution and field performance of a rescue robot', *Journal of Field Robotics*, **25**(1–2), 17–30, (2008).

[30] N. Moray, T. Inagaki, and M. Itoh, 'Adaptive automation, trust, and self-confidence in fault management of time-critical tasks.', *J Exp Psychol Appl*, **6**(1), 44–58, (2000).

[31] B. Muir, 'Trust between humans and machines, and the design of decision aids', *International Journal of Man-Machine Studies*, **27**(5–6), 527–539, (1987).

[32] T. Nomura, T. Kanda, T. Suzuki, and K. Kato, 'Prediction of Human Behavior in Human–Robot Interaction Using Psychological Scales for Anxiety and Negative Attitudes Toward Robots', *IEEE Transactions on Robotics*, **24**(2), 442–451, (2008).

[33] R. Parasuraman and C. Miller, 'Trust and etiquette in high-criticality automated systems', *Communications of the Association for Computing Machinery*, **47**, 51–55, (2004).

[34] R. Parasuraman and V. Riley, 'Humans and automation: use, misuse, disuse, abuse', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, **39**(2), 230–253, (1997).

[35] R. Parasuraman, TB Sheridan, and CD Wickens, 'A model for types and levels of human interaction with automation', *Systems, Man and Cybernetics, Part A, IEEE Transactions on*, **30**(3), 286–297, (2000).

[36] M. Rehak, L. Foltyn, M. Pechoucek, and P. Benda, 'Trust model for open ubiquitous agent systems', in *Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT'05)*, (2005).

[37] V. Riley, 'Operator reliance on automation: theory and data', in *Automation and Human Performance: Theory and Applications*, (1996).

[38] V. Riley, 'Operator reliance on automation: Theory and data', *Automation and human performance: Theory and applications*, 19–35, (1996).

[39] J. Sanchez, *Factors that affect trust and reliance on an automated aid*, Ph.D. dissertation, Georgia Institute of Technology, May 2006.

[40] W. Song, V. V. Phoha, and X. Xu, 'An adaptive recommendation trust model in multiagent systems', in *Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT'04)*, (2004).

[41] Transportation Research Associates, 'Project no. 99-06: Intelligent vehicle initiative (side collision warning system) operational test evaluation'. Final Evaluation Report (Revised), 2003.

[42] Mechanical Turk. `http://www.mturk.com`.

[43] A. Uggirala, A.K. Gramopadhye, B.J. Melloy, and J.E. Toler, 'Measurement of trust in complex and dynamic systems using a quantitative approach', *International Journal of Industrial Ergonomics*, **34**(3), 175–186, (2004).

[44] S. van Mulken, E. Andre, and J. Muller, 'An empirical study on the trustworthiness of life-like interface agents', in *Proceedings of the HCI International*, volume 99, pp. 152–156, (1999).

[45] H.A. Yanco, J.L. Drury, and J. Scholtz, 'Beyond Usability Evaluation: Analysis of Human-Robot Interaction at a Major Robotics Competition', *Human-Computer Interaction*, **19**(1&2), 117–149, (2004).

Table 3. A list of papers that examine factors which influence trust in automation.

Paper	Task/Automation type	Number of Participants	Participant Background	System Type	Significant Findings
Lee and Moray 1991 [19]	Orange juice pasteurization plant	12	Students	Simulation	The time that users spend using automation depends not only on trust and self-confidence but also on their past use of automation and their individual biases.
Riley 1996 [38]	Character identification	30	Students and pilots	Simulation	Among other things, their data did not reveal evidence of positivity bias over the full stimulus experience. In addition, they compared the automation curves of expert and non-expert users (pilots and students). The pilots' responses followed the automation allocation curve of students except that the pilots showed a greater bias towards automation.
Jian et al. 1998 [16]	Word elicitation	157	Students	N/A	Trust and distrust can be treated as opposites, lying along a single dimension of trust.
van Mulken et al. 1999 [44]	Multimedia presentation system	32	Not given	Real	The researchers found no statistical difference between the trustworthiness of personified and unpersonified agents.
Herlocker et at. 2000 [15]	Automated collaborative filtering (ACF)	210	Volunteer users	Real	The researchers seek to find ways to improve performance of an expert system through different explanation modalities and find that, in general, adding an explanation interface to an ACF system improves the acceptance of that system among users. Results were inconclusive.
Moray et al. 2000 [30]	Central heating system	30	Students	Simulation	The researchers answer several questions about adaptive allocation and trust with respect to time-critical systems and faults.
Dzindolet et al. 2003 [9]	Camouflaged soldier detection	219	Students	Simulation	This work provides support for the positivity bias. The results also indicated that, after considerable interaction with the automated decision aid (ADA), the operators were less forgiving of automation errors. However, by providing explanations of why the ADA might make a mistake increased the operators' trust in the ADA.
Khasawneh et al. 2003 [18]	Hybrid inspection system	12	Students	Simulation	Trust is proportional to the performance of the system and can also be predicted based on system errors.
Uggirala et al. 2004 [43]	Line length estimation	12	Students	Simulation	The authors use uncertainty as a quantitative alternative to trust and find that overall trust in a system is inversely correlated with system uncertainty.
Antifakos et al. 2005 [1]	Context aware systems	14	Students	Simulation	Explicitly displaying the current confidence of the system increases the usability of automatic / context-aware systems.
Madhavan et al. 2006 [24]	Visual inspection	45	Students	Simulation	The easiness of errors made by decision aids affects users' trust.
Masthoff 2007 [26]	Perception of trust in a person given a written scenario about that person	49	Volunteers from CS Dept.	N/A	The paper introduces a quantitative trust model that includes factors not generally considered by many researchers, such as direct experiences, stereotypes, reputation, empathy, and user characteristics.
Glass et al. 2008 [14]	Adaptive agent system	14	Employees	Real	The researchers identify and discuss eight major themes that significantly impact user trust in complex systems.
Merritt et al. 2008 [28]	X-ray screening task	255	Students	Simulation	Individual differences account for a large variation in trust when system characteristics are kept constant.

Haptic Control for the Interactive Behavior Operated Shopping Trolley InBOT

Michael Göller and **Thilo Kerscher** and **Marco Ziegenmeyer** and **Arne Rönnau** and **J.Marius Zöllner** and **Rüdiger Dillmann** [1]

Abstract. An important challenge in service robotics is to design user interfaces that can be used intuitively to enable the potential users to benefit from the robot's functionalities. These interfaces have to be easy to understand on the one hand but have to be able to access the various features that are provided by robotics. This document presents the haptic-based control of an Interactive Behavior Operated shopping Trolley (InBOT). The holonomic robot InBOT is an approach to transfer state of the art robotics technology into all-day environments to make their benefits available for the populace. The haptic-based mode of control enables the user to steer InBOT just like an ordinary shopping cart while being released from the burden to push the cart with muscle power. Additionally haptic commands are introduced which the user can give the robot using the haptic handle. These commands are classified by a vector support machine and activate corresponding behaviors of the robot's advanced behavior repertoire. Finally the user is supported continuously by features from the robot's basic behavior repertoire like the avoidance of obstacles.

1 Introduction

The Interactive Behavior Operated Trolley (InBOT, fig.1) is an approach to transfer state of the art robotics technology into human everyday environments. The benefits of the actual technology shall become available for the populace. InBOT addresses several everyday problems and its main idea is to ease the daily shopping. Among other possibilities this means to help the customer to find the desired products without extensive search in big supermarkets or to relieve the customer from the burden to push the shopping cart from his own force all the time, especially if the cart is heavily loaded or the customer is an elderly person.

InBOT provides four different modes of operation. Their interaction with the remaining system is roughly sketched in fig.2. The first one is the *haptic steering mode*. This is the mode with the closest coupling between user and robot. Here InBOT is pushed like an ordinary shopping cart but supports the user with several features like the full motor power of the robot, an obstacle avoidance assistant or local haptic commands. The transition into this mode is special. While for the other modes an explicit command of the user is necessary InBOT automatically switches to the *haptic steering mode* as soon as a force is applied to the handle. The second one is the *following mode*. Here the robot follows the user wherever he goes in a well-defined distance. The third one is the *guiding mode*. In this mode the robot guides the user to a desired target or product or even

Figure 1. The Interactive Behavior Operated shopping Trolley (InBOT). It is equipped with a mecanum drive for full holonomic movements and a haptic handle to enable the user to steer it just like an ordinary trolley.

along a whole shopping list step by step while continuously adapting to the user's movements. The last mode is the *autonomous mode* with the weakest coupling between user and robot. Here InBOT acts only if explicitly commanded by the user and once commanded the robot performs the task independently. As baseline of all these modes acts a *behavior-based control* ([1]) that provides a repertoire of behaviors to be used in the course of the four modes. Among the provided behaviors is for example the capability to localize itself and to find a way to a desired product or the avoidance of obstacles or dynamic objects. This document describes the haptic control of InBOT which is the most crucial part of the *haptic steering mode*.

The most intuitive method of using the robot trolley is probably the interaction based on physical contact by using a haptic handle (fig.3). The user doesn't need extensive training and doesn't have to learn voice or gesture commands. Additionally this kind of interaction has a high degree of reliability. The close coupling of user and robot in the control task allows the user to perceive the robot's actions while allowing him to contribute his own intentions simultaneously. Thus he preserves a feeling of control and responsibility ([3]). Due to these advantages the *haptic steering mode* might be the most important mode of controlling InBOT. The user is able to apply ex-

[1] FZI Research Center for Information Technology, Karlsruhe, Germany, email: [goeller, kerscher, ziegenmeyer, roennau, zoellner, dillmann]@fzi.de

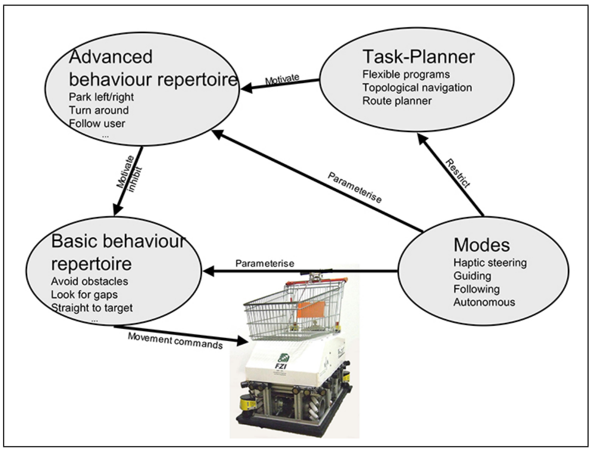


Figure 2. The abilities of InBOT are grouped in 2 repertoires of behaviors with a flexible task-planner on top. The different modes of operation influence the work of the different behaviors and of the task planner([2]).

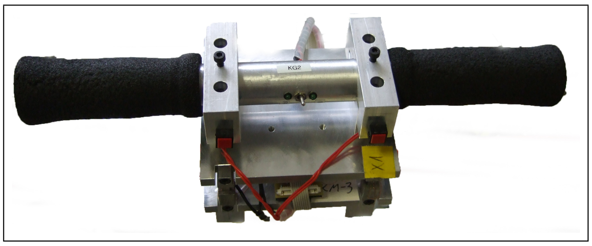


Figure 3. Close-up of the haptic handle with the two red dead man's switches.

periences from a common, manual trolley to the robot trolley but is supported by several additional functionalities like motor power, obstacle avoidance or local maneuvers controlled by haptic commands.

2 Control chain for the haptic handle

To achieve a high level of usability the trolley's handle must be able to detect the applied forces and torques, gather them and transfer them to a classification mechanism that is based on a *support vector machine*(SVM). After the classification motion commands are generated and transferred to the behavior network ([1]) implementing the *basic behavior repertoire*. After the movement commands have been merged with the commands from other behaviors the final movement command is handed down to the motion control in the hardware abstraction layer.

In detail (see fig.4) this means the following: The forces that the user exerts on the bars of the haptic handle are measured by strain gauges. The values are digitalized and transferred to a UCOM-board ([4]). Here the several measured forces are transformed into two forces and one torque, which are the most significant information used in the haptic mode. This procedure of data acquisition is described in detail in the next section. Additional signals are provided by two buttons on the handle that are used as dead man's switches. The UCOM-board transfers the processed forces to the navigation system only if at least one of the buttons is pressed. The second functionality of the buttons is to trigger the classification of haptic commands. At the moment when both buttons are released the responsible *support vector machine* processes the forces measured in the last second and classifies the given command (see section 2.3 for a detailed description). Each identified command motivates one of the behaviors of the *advanced behavior repertoire*. While at least one of the buttons is pressed the actual command is classified as *haptic steering* and the processed forces are transformed into movement commands (section 2.2). Based upon the classification and the calculated information the different behaviors of the behavior repertoires can be activated.

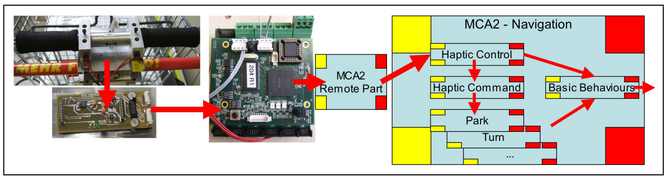


Figure 4. Control chain of the haptic mode: Haptic handle - ACAM-Board - UCOM-Board - MCA2 Remote part - Module Haptic Control - Module Haptic Command Classification (based on SVM) - Advanced Behaviors - Basic Behaviors

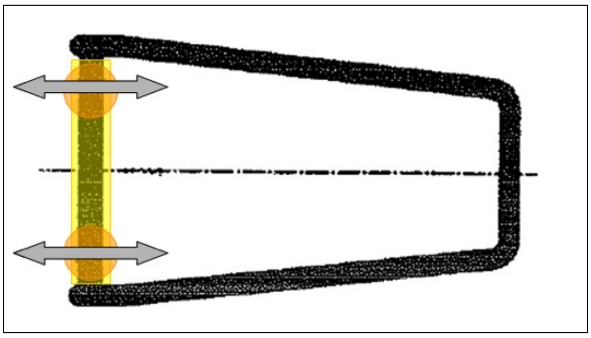


Figure 5. Typical forces applied by human users.

2.1 Haptic-based detection of commands

To sense and detect the user-intention using the trolley's handle a physical contact of user and trolley is necessary. The user grabs the handle at least with one hand. The possible area for the physical contact of the user is illustrated in fig.5 by the yellow rectangle. The most typical areas for the contact are illustrated by the orange circles. The arrows demonstrate the most typical direction of the forces applied by the user.

The user is able to apply forces as well as torques to the trolley's handle one or both hands. During contact of both hands the rigid body of the handle will combine the different forces and torques to a resulting force and torque (see fig.6). For the steering, but also for the classification of the intention, only the projection of the force on the horizontal plan (fig.7) as well as the torque around an axis parallel to the normal vector of the horizontal plan is used.

The sensorized trolley handle must detect three values which are the absolute value of the force, the direction of the force, and the torque around the normal vector of the horizontal plane. The two values for the force should be orthogonal.

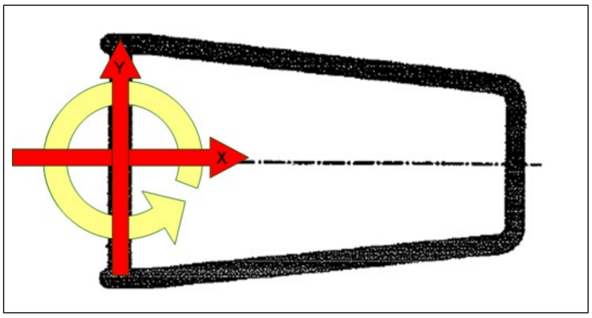


Figure 6. Forces (red arrows) and torque (yellow circle) that are finally provided by the haptic handle.

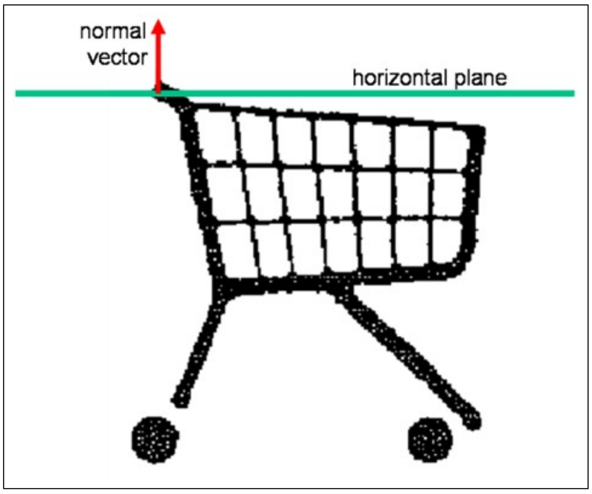


Figure 7. Horizontal plane and normal vector for the force and torque direction

2.1.1 *Mechanical design of the haptic handle*

The current setup of the haptic handle focuses on the elementary requirements which have been described above. This setup can be used as a haptic handle for the robot trolley as well as a sensor of a common trolley to gather sensor information for example for the user-centered prototyping evaluation. The setup of the haptic handle is shown in fig.3.

The main part which hosts the force and torque sensor consists of two planes, which are connected by four strain gauges beams. One plane is directly connected to the trolley. The second plane carries the handle, which will be used by the user to steer the trolley. Due to this setup the flow of force must go from the handle through the second plane, the strain gauges beams and finally the first plane to the trolley.

The strain gauges for the detection of the forces and the torques are placed on three of the four strain gauges beams. Figure 8 illustrates the positions of the different strain gauges and the force direction they sense. The beams B1 and B2 are equipped with strain gauges to sense the forces in the x-direction. The beams B1 and B3 are equipped with strain gauges to sense the forces in the y-direction. The number of four strain gauges is sufficient because in general three forces (with independent force directions) will be enough to calculate the forces and the torque of the haptic handle.

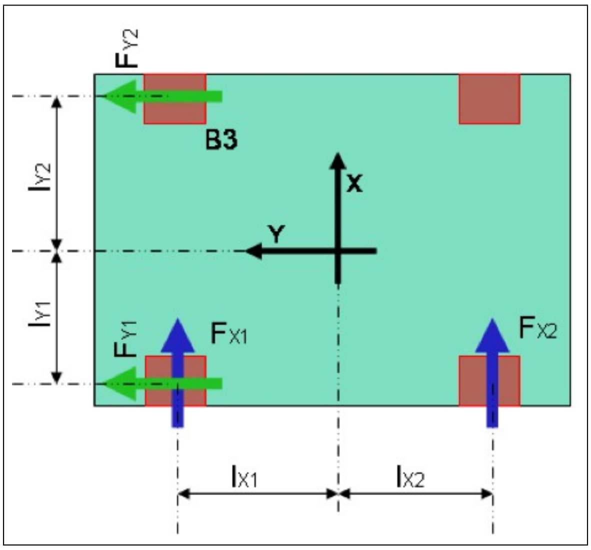


Figure 8. Base of the haptic handle with indicated resulting force directions (black arrows). The green arrows indicate the two different forces measured in Y-direction, the two blue arrows indicate the forces measured in X-direction.

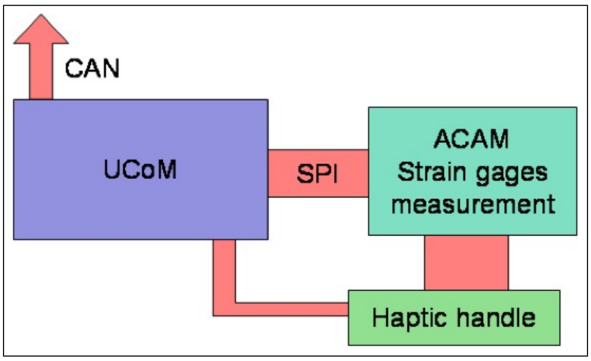


Figure 9. Schematic diagram on the electronic components of the haptic handle

2.1.2 *Electronic components of the haptic handle*

The electronic components of the haptic handle are responsible for the measurement of the strain gauges, the measurement of the human contact, the pre-processing of the gathered sensor information as well as the communication with the overall system. Figure 9 illustrates the used components and their connections.

For the measurement of the strain gauges the specialized component ”PS021 - Digital Amplifier for Strain Gauges” from acam-messelectronic GmbH (www.acam.de) is used. The main advantage of this component is that the integrated circuits are realized without any analogue component but an approved Time-to-Digital Converter (TDC) technology. This technology has a lot of advantages like there is no need for a full-bridge, 2 resistors (half-bridge) are sufficient or no reference voltage is needed. The PS021is connected to the UCoM over SPI-BUS.

The major part of the components is the Universal Controller Module (UCoM), which is responsible for the gathering as well for the pre-processing of the sensor information. This board has been developed at the FZI as a modular hardware component for example for lower level control tasks (motor control) as well as sensor preprocessing. A detailed description of the UCoM can be found in [4]. The UCoM is encapsulated in a network-transparent remote part using the MCA2 framework ([8], [9]) and communicates with the overall system over CAN-BUS. The UCoM is directly connected to the sensors for the contact of the user.

2.1.3 Calculation of the applied forces and torques at the haptic handle

The major part of the software component is the calculation of the applied forces and torques at the haptic handle. The positions of the measured forces are illustrated in fig.8. The different measured forces are not independent enough to use them as raw data for the calculation. This is caused by manufacturing tolerances, varying material properties and the fact that the manual placed strain gauges are not total identically positioned. So a calibration of the total assembled system is necessary. The calibration is done by linear calibration using a 4x4 calibration matrix **A** which transfers the raw sensor values $\vec{RS} = (F_{XR1} F_{XR2} F_{YR1} F_{YR2})^T$ to the calibrated ones $\vec{CS} = (F_{X1} F_{X2} F_{Y1} F_{Y2})^T$ using the relation:

$$\vec{CS} = \mathbf{A} \cdot \vec{RS} \tag{1}$$

The calibration matrix transfers the four raw values to more independent values. So afterwards the force values can be treated as independent. The calibration matrix has been determined by applying the least square method to calculate the necessary elements of the calibration matrix. For this purpose a defined set of independent test forces has been applied on the haptic handle.

Now, the resulting force and torques can be calculated with the following equations:

$$F_X = F_{X1} + F_{X2} \tag{2}$$

$$F_Y = F_{Y1} + F_{Y2} \tag{3}$$

$$T = -F_{X1} \cdot l_{X1} + F_{X2} \cdot l_{X2} - F_{Y1} \cdot l_{Y1} + F_{Y2} \cdot l_{Y2} \tag{4}$$

2.2 Generation of haptic-based movement

If the module responsible for the haptic control receives forces from the UCOM-board it demands a mode change to the haptic mode at once. This request is processed in a higher layer. If operating in the haptic mode the module transforms the measured forces into a velocity vector using a dynamic approach. Formula 5 describes how the set point vector for the velocity (v_t) is calculated. It depends on the past velocity, the proportion of the measured force (F) compared to the maximum accepted force (F_{max}), an amplification factor (f_a) and a damping factor (f_d). The state variable v represents the set point vector for the desired velocity. It is amplified by the measured forces and damped by its actual value. This velocity vector is then passed to the behavior-based control. Here it is merged with velocity vectors of other behaviors like *obstacle avoidance* and finally passed to the hardware abstraction layer. Here the resulting desired velocities are transformed in engine speed values for the mecanum drive.

$$v_t = v_{t-1} + min\left(\frac{|F|}{F_{max}}, 1\right) \cdot sign(F) \cdot f_a \cdot \Delta t - f_d \cdot v_{t-1} \cdot \Delta t \tag{5}$$

2.3 SVM-based classification of haptic commands

Another alternative to generating velocity vectors out of the measured forces directly is to analyze the measured forces over a period of time and to try to identify haptic commands. If identified these commands overrule the direct generation of velocity vectors. An identified command activates the corresponding behaviors from the *advanced behavior repertoire*. The activated advanced behaviors are responsible for the generation of the velocity vectors themselves. Usually they again activate behaviors from the *basic behavior repertoire* (see fig.4) .

The discrimination between these two approaches is realized by using the haptic handle's contact sensors. If a user has contact with the haptic handle the first approach is active. If the user releases the handle the past temporal progression of the applied forces and torques is used to classify the different intentions of the user.

2.3.1 Classes of haptic commands

The definition of the classes is motivated by the different user-intentions which should be distinguisted by the classificatory. Each of this intentions is connected to a more haptic handle specific description, which is the basis of the command classes. The following intentions of the user are supported: ”Wait exactly here”, ”Stay near the actual position”, ”Rotate to support loading”, ”Park at side”, ”Drive a little forward”.

- ”Steer”
  - Continuous contact between user and handle
  - → Class name: STEER
- ”Wait exactly here”
  - Handle has been released
  - During release a high force is applied on the handle in backward direction
  - → Class name: PULL
- ”Stay near the actual position”
  - Handle has been released
  - At time of release no force and torque has been applied on the handle
  - → Class name: WAIT
- ”Rotate to support loading”
  - Handle has been released
  - During release a high torque is applied on the handle
  - → Class name: ROTATE LEFT
  - → Class name: ROTATE RIGHT
- ”Park at side”
  - Handle has been released
  - During release a high force is applied on the handle
  - → Class name: PUSH LEFT
  - → Class name: PUSH LEFT
- ”Drive a little forward”
  - Handle has been released
  - During release a high force is applied on the handle
  - → Class name: PUSH

The haptic commands which result from the classification of the user intention are used to command the trolley. The resulting classes of commands are summarized in table 1.

Table 1. Supported classes, the corresponding force characteristics and the corresponding advanced behavior that is triggered after the classification.

Class name	R	FX	FY	T	AB
STEER	No	-	-	-	Steer
PULL	Yes	H(-)	S	S	Stay here
WAIT	Yes	S	S	S	Stay in area
ROTATE L/R	Yes	S	S	H(+/-)	Support loading
PUSH L/R	Yes	S	H(+/-)	S	Park
PUSH FRONT	Yes	H(+)	S	Sl	Drive forward

R: Release of handle
F: Value of force in [X,Y]-direction. Values: High(sign) or Small
T: Value of torque. Values: High(sign) or Small
AB: Corresponding advanced behavior

The relevant information is the state of the dead man's switches (pressed/released) and the two measured forces and the torque. The last column names the corresponding advanced behaviors that are activated after identifying a command.

2.3.2 *Definition of feature classes as basis for the classification*

For the quality of the classification the definition of significant feature classes is mandatory. For example taking only the raw force and torque progressions of different measurements with different strength would not be generalized to one class even if the intentions was identical.

To gain the best result 23 feature classes have been defined. Some of them have been preliminary defined like the average value for the different progressions. Other feature classes have been defined during the learning and evaluation to improve the quality of the classification like for example the amount of positive force values or the dominant force direction: If the absolute value of force x is bigger than the absolute value of force y then $dom(F_X) = \pm 1$. The algebraic sign is determined by comparing the value of min and max. If min is larger then the sign is negative. Finally the used feature classes are reported in the table 2. For the usability with the SVM it is necessary that all input data is standardized.

Most of the features cannot clearly be assigned to one of the classes directly. Some features are irrelevant for some classes. The method of how the class "ambiguous" is assigned is very important for the classification system in general. Naturally for all implementations of artificial intelligence there are ambiguous decisions, just like in real life. For SVM this ambiguity is manifested in different ways. If the learning has been incomplete no activations for any class are present. In this case as well as for activations that are below a specified threshold the class "ambiguous" is assigned. These assignments represent incomplete training, which is of course also possible for rather matured learning machines. If, to the contrary, the learning has advanced very far, multiple activations are possible. In such a case a rule can be conceived of that would tell what class activations overrule what other class activation. It was deemed clearer and for the beginning to also classify these features as ambiguous.

2.3.3 *Support vector machine for the haptic-based classification of user intention*

As input for the classification the different sensor signals coming from the haptic handle are used. These are the forces in X- and Y-direction, the torque and the contact between user and trolley. These information are accumulated over a period of time. The time progression of these sensor values is analyzed. Crucial for the success of the learning and later classification of the different classes is the calculation of distinctive features which was described in the prior section. These features make up the vectors for the *Support Vector Machine* (SVM).

Table 2. Defined features for the classification

Feature class	Description	Range
$avr(F_X)$ $avr(F_Y)$ $avr(F_T)$	Average value for force/torque in the period	[-1,...,1]
$min(F_X)$ $min(F_Y)$ $min(F_T)$	Minimum value for force/torque in the period	[-1,...,1]
$max(F_X)$ $max(F_Y)$ $max(F_T)$	Maximum value for force/torque in the period	[-1,...,1]
$\#neg(F_X)$ $\#neg(F_Y)$ $\#neg(F_T)$	Number of negative values for force/torque in the period	[0,...,1]
$\#pos(F_X)$ $\#pos(F_Y)$ $\#pos(F_T)$	Number of positive values for force/torque in the period	[0,...,1]
$max(diff\|F_X\|)$ $max(diff\|F_Y\|)$ $max(diff\|F_T\|)$	Maximum value of the absolute value of the first derivation	[-1,...,1]
$dev(F_X)$ $dev(F_Y)$ $dev(F_T)$	Standard deviation for force/torque in the period	[0,...,1]
$dom(F_X)$ $dom(F_Y)$	Investigation of dominant force direction	[-1,0,1]

$F_{...}$: Value of force in [X,Y]-direction
F_T: Value of torque.

SVMs belong to the relatively new family of Kernel Methods, that combine the simplicity and computational efficiency of linear algorithms, such as the perceptron algorithm, with the flexibility of non-linear systems, such as for example neural networks, and the rigor of statistical approaches such as regularization methods in multivariate statistics ([5]). By reducing the learning step to a convex optimization problem, which can always be solved in polynomial time, the problem of local minima typical of neural networks, decision trees and other non-linear approaches is avoided. Therefore the training of support vector machines is deterministic and retraining is faster and easier.

Moreover, due to their foundation in the principles of Statistical Learning Theory ([6]) they are remarkably resistant to over fitting especially in circumstances where other methods are affected by dimensionality.

The absolute value of the resulting decision function is called activation. It corresponds to the distance between the projected data input and the separating hyper plane in the vector space. Therefore it indicates the classification quality for a given feature: a low activation stands for an unconfident classification and a high activation for a confident classification ([7]).

3 Experimental results

Several experiments were performed to evaluate the *haptic steering* mode for InBOT. Figure 10 shows a user pushing InBOT like an ordinary shopping cart. He is supported by various features. First of all no muscle power is required to move the robot even if heavily loaded. Second InBOT automatically avoids collisions with obstacles. The user is even able to grab the handle with one hand only like depicted in figure 11. Even if the user doesn't recognize the obstacle himself InBOT moves around it safely. The effect of the responsible behaviors from the *basic behavior repertoire* is depicted in figure 12. Additionally the user is able to control InBOT by using haptic commands that are classified by a SVM.

Figure 10. The user is steering InBOT like an ordinary shopping cart using the handle. He is supported by InBOT's motors.

To maximize the outcome and the quality of the classification it is important to have significant and independent data for each class. For the final test on the learning of the developed classification 4 test persons developed learning data. Each test person has contributed ten measurements per class. In the end a total number of 280 tests were collected. Each class has 40 measurements. For the verification the data of one test person has been used. Here each class has 20 measurements, in total there have been 140 tests for evaluation.

Figures 13 to 18 show representative measurements together with the identified class to get an impression about the overall outcome of the sensor data. Figure 19 shows InBOT reacting the command ROTATELEFT.

Using overall 280 tests of several different users the SVM has been trained. All data has been verified as valid for testing. As result of the learning 86 support vectors have been used to generalize the collected data. For verification of the learned classes a set of 140 tests has been used. 136 measurements have been classified correct. The classification of 4 measurements has not been successful. So the error rate is 2,8 %. Looking at the confusion matrix one can see that the classes WAIT, PUSH, PUSH LEFT, PUSH RIGHT and PULL have been classified correct (100%). For that set of data only for the two classes ROTATE LEFT and ROTATE RIGHT a full correct classification was not possible. For the class ROTATE LEFT one measurement has been classified as PUSH LEFT. For the class ROTATERIGHT three measurements have been not correctly classified, one of them has been classified as PUSH RIGHT the others have not been classified at all. The misclassification of the rotational classes and the sideward-push classes is due to the similar forces in the time progression. The reason for this is that it is difficult for the human user to push or rotate the trolley without implicitly applying additional forces or torques. The details of the confusion matrix are shown in table 3.

Figure 11. The user is steering InBOT like an ordinary shopping cart using the handle. Because he needs his left hand otherwise he grabs the handle with his right hand only. This makes it difficult for him to maneuver dexterously. So the user benefits from being supported by by the *obstacle avoidance* of the *basic bahavior repertoire*.

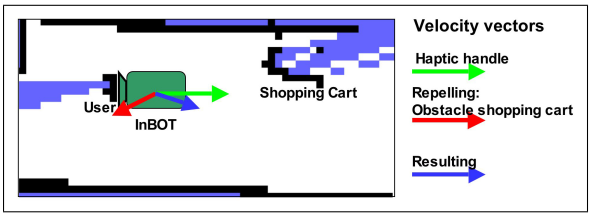


Figure 12. Occupancy grid of the scene from figure 11. The influence of the supporting *avoid obstacle* behaviors is illustrated by the different arrows ([1]).

Table 3. Confusion matrix: percentage of correctly classified commands. The rows represent the commands given by the user, the columns the result of the classification. The values represent the percentage of how often one command was assigned to a specific class by the SVM.

	WAIT	PUSHF	PUSHL	PUSHR	PULL	ROTL	ROTR	NONE
WAIT	1,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
PUSHFRONT	0,0	1,0	0,0	0,0	0,0	0,0	0,0	0,0
PUSHLEFT	0,0	0,0	1,0	0,0	0,0	0,0	0,0	0,0
PUSHRIRGHT	0,0	0,0	0,0	1,0	0,0	0,0	0,0	0,0
PULL	0,0	0,0	0,0	0,0	1,0	0,0	0,0	0,0
ROTATELEFT	0,0	0,0	0,0	0,05	0,0	0,95	0,0	0,0
ROTATERIGHT	0,0	0,0	0,05	0,0	0,0	0,0	0,85	0,1

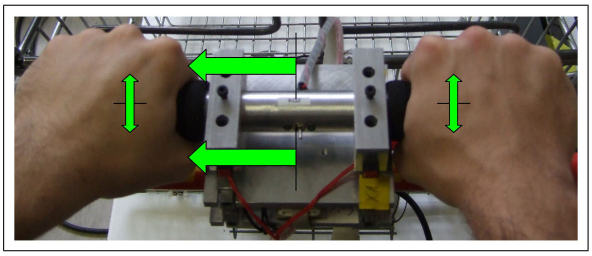


Figure 13. The haptic handle in use with the command PUSHLEFT indicated by the green arrows. Large forces are exerted in Y-direction to the left, small forces are exerted in X-direction.

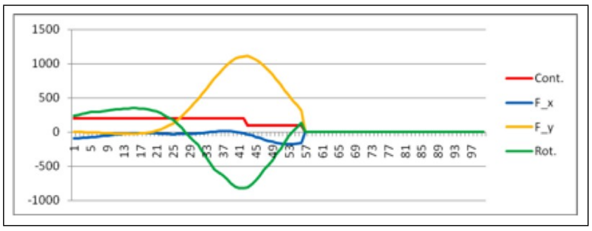


Figure 14. Example for measured forces that were classified as PUSHLEFT.

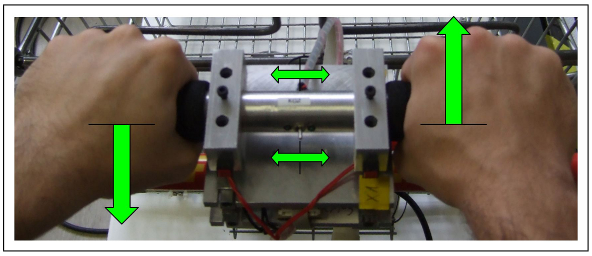


Figure 15. The haptic handle in use with the command ROTATELEFT indicated by the green arrows. Large forces are exerted in positive (right side) and negative (left side) X-direction, hardly and forces are exerted in Y-direction.

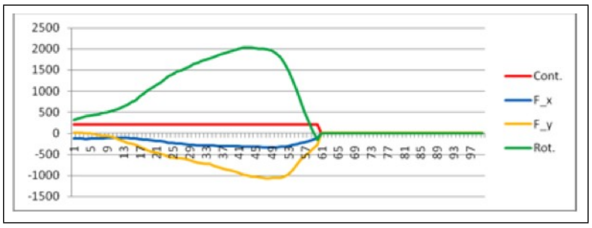


Figure 16. Example for measured forces that were classified as ROTATELEFT.

4 Conclusion

For the intuitive use of the trolley a haptic interface has been developed. Based on the physical contact of user and trolley different user intention have been derived such as a ”push” command or a ”steering” command where user and robot move together using a version of zero-force-control. This haptic interface is realized using a sensorized trolley handle. This handle is equipped with different sensors for detecting the forces and torques applied by the user.

Based on the gathered sensor data a classification using *support*

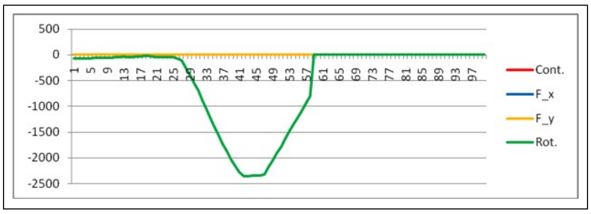


Figure 17. Example for measured forces that were classified as ROTATERIGHT.

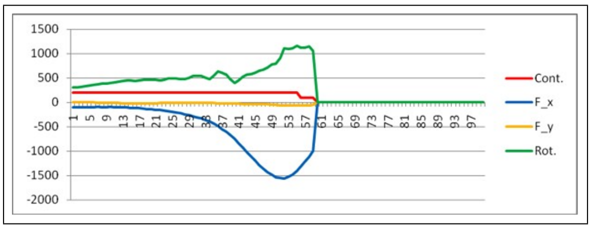


Figure 18. Example for measured forces that were classified as PULL.

Figure 19. Example of ROTATELEFT: The user pushes InBOT to a heavy crate and gives the command ROTATELEFT to ease the loading (left). InBOT turns around and stops with the side to the user (right).

vector machines has been developed, implemented and tested. For the classification the following classes have been defined: ”Steer”, ”Wait here”, ”push left”, ”push right”, ”push forward”, ”push backward” and ”Turn around”. A major concern was the discrimination of steering and the other commands to avoid possible confusions of the user. If user and trolley have contact always the steering command is active. Therefore a direct control of the robot InBOT has been developed. Only if the user releases the handle the classificator for the other command classes is used. This classificator has been trained using collected sensor data of various people (280 measurements) and then verified (140 measurements) with a success rate of 97,1%. The identified classes are then used to activate advanced behaviors. Further user studies will be performed in the future to determine the necessity of each different command from the user’s point of view. A possible reduction of the number of commands will probably lower the confusion even more. This is very important even if the classification performs very well because each false-positive results in an unpredictable movement. On the other hand the user studies could reveal the need of one or more additional commands.

From the author's point of view the haptic control of a robot shopping cart is very crucial because it is very intuitive. This way of control most untrained users will try first simply because they are used to use a shopping cart this way. Additionally the user has the feeling of really controlling the robot, not only of being a spectator. Therefore the authors think that it was important to develop such an intuitive possibility of robot-control to raise the level of acceptance for this new technology. The main idea of InBOT was to make robotic technology available in all-day situations, so it was obvious not only to develop the haptic steering of the robot trolley, but also to enhance it with augmentations known from robotics like the obstacle avoidance or to give the user additional haptic-based tools to control the robot. Some of these are even known to customers, for example pushing the (empty) conventional trolley sideways to park it at a side of the corridor, others on the other hand are hopefully a new enrichment of the shopping experience.

ACKNOWLEDGEMENTS

This research has been carried out in the CommRob project (www.commrob.eu) and is partially funded by the EU (contract number IST-045441 under 6th framework programme).

REFERENCES

[1] M.Göller, T.Kerscher, J.M.Zöllner, R.Dillmann: Behavior Network Control for a Holonomic Mobile Robot in Realistic Environments, CLAWAR 2008

[2] M.Göller, F.Steinhardt, T.Kerscher, J.M.Zöllner, R.Dillmann: RFID Transponder Barriers as Artificial Landmarks for the Semantic Navigation of Autonomous Robots, CLAWAR 2008

[3] Paul Griffiths and R. Brent Gillespie: Sharing Control Between Human and Automation Using Haptic Interface: Primary and Secondary Task Performance Benefits, Human Factors: The Journal of the Human Factors and Ergonomics Society, Vol. 47, No. 3, Fall 2005, pp. 574-590.

[4] K. Regenstein, T. Kerscher, C. Birkenhofer, T. Asfour, J.M. Zöllner, R. Dillmann: Universal Controller Module (UCoM) - component of a modular concept in robotic systems. IEEE International Symposium on Industrial Electronics, 2007.

[5] N. Cristianini and J. Shawe-Taylor, 2003, "Support Vector and Kernel Methods", In: Berthold, Michael and David J. Hands (Eds.): Intelligent Data Analysis. Springer-Verlag Berlin Heidelberg New York.

[6] V.N. Vapnik, "Statistical Learning Theory", 1998, John Wiley & Sons.

[7] H. Willems, K. Reber, J. Zöllner, M. Ziegenmeyer: New Approaches to the automated analysis of ultrasonic in-line inspection data. ASME IPC 2004, Calgary, Canada.

[8] K. U. Scholl, J. Albiez, and B. Gassmann, "MCA - An Expandable Modular Controller Architecture," in 3rd Real-Time Linux Workshop, 2001, Milano, Italy, 2001.

[9] FZI, "Modular Controller Architecture Version 2." [Online]. Available: http://www.mca2.org/

Photogeometric Sensing for Mobile Robot Control and Visualisation Tasks

Alonzo Kelly[1], Dean Anderson[1], Erin Capstick[2], Herman Herman[1], Pete Rander[1]

Abstract. Photogeometric sensing is a relatively new sensor modality that tightly integrates geometry and appearance sensing into a single package. Such a sensor produces imagery that encodes the appearance and the range to every sensed point in the scene. This new type of sensor enables much higher fidelity virtualized reality displays that can be produced in real time from the data gathered by a moving robot. Such displays exhibit several ideal characteristics for human robot interaction tasks that enable new approaches to supervisory control and remote visualization. Photogeometric sensors suitable for HRI applications cannot yet be purchased but they can be constructed by co-locating ranging and appearance sensors and combining the data at the pixel level. This paper outlines our approach to the construction of such sensors as well as their successful use in several applications.

1 INTRODUCTION

We will use the term *appearance* to refer to sensing modalities which are sensitive to the intensity of incident radiation including visible color, visible intensity, and infrared modalities. Conversely, *geometry* will be used to refer to modalities that register any of depth, range, shape, disparity, parallax, etc. The term *photogeometric* (PG) sensor will refer to a sensing device that produces both kinds of data in a deeply integrated manner. For our purpose in this paper, the data is deeply integrated if the spatial correspondences of the data are known. Ideally, as shown in Figure 1, the resolutions are matched as well so that a one-to-one mapping exists between geometry and appearance .

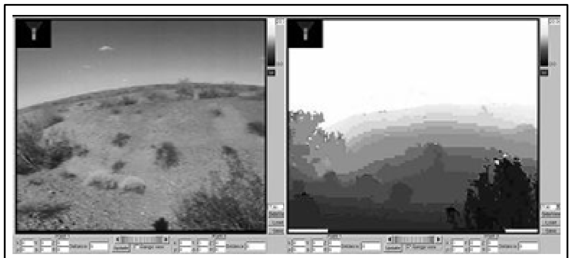


Figure 1: Photogeometric Data Set. Every color pixel in the left image has an associated range pixel in the right image.

The deep integration of appearance and geometry data can be a powerful technique for enabling effective human-robot interaction. In many applications, robots must understand the geometry of the environment in order to move around competently while avoiding collision. In such applications, geometry sensing is often the preferred modality of the robot designers. Conversely, humans process appearance data more readily and we can assimilate geometry perceptually only when it is converted to appearance data. For example, two stereo views or the parallax evident in a moving cloud of points on a computer screen will enable humans to perceive depth.

When images of both modalities are available – and their correspondence is known – it becomes possible to convert between the modalities relatively seamlessly. For supervisory control, such conversion makes it possible to extract accurate 3D coordinates when a pixel in a video stream is designated. For visualization, such conversion makes it possible to render synthetic views of the scene from arbitrary perspectives which may never have been the site of any real sensor.

The paper is organized as follows. Section 2 provides a broad overview of related work. Section 3 explains our technique for producing a photogeometric sensor. Section 4 describes an experiment using such sensing for mobile manipulator teleoperation. Section 5 describes an experiment using such sensing for outdoor mobile robot teleoperation. Section 6 provides a brief summary and outlook.

2 RELATED WORK

The notion of aiding an operator by displaying the perception data produced by a remotely controlled robotic system must have occurred to the first designers of such systems. Numerous techniques for supervisory control and teleoperation of manipulators, and even telepresence were clearly outlined as early as the mid 1980s [16]. The same concepts were considered early for legged vehicles [13] and wheeled mars rovers [3].

In broad terms, although perception data is nominally a view of the state of the environment, it is more properly described as a view of the robot's model of that environment. Hence, such data is equally a view into the internal state of the robot. It is natural for engineering displays to use such data during system development but it also quickly becomes clear that a good way to understand robot behavior is to know what it "thinks" it perceives in its immediate surroundings.

Of course, this mode of tele-operation depends on the use of adequate sensing. Military and consumer markets have driven the development of guidance systems and TV cameras that are relevant to mobile robots today. While laser ranging sensors are now commercially produced for factory robots, systems designed specifically for outdoor mobile robots are either single axis, immature products, or of inadequate performance for our purposes. For these reasons, our work continues a long tradition [11] of custom sensor development for lack of any alternative

[1] Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA, 15213
[2] DCS Corporation. 1330 Braddock Place, Alexandria, VA, USA, 22314

which continues in robotics labs around the world up to the present time [12].

Given the sensor data needed, the earliest approaches simply displayed the raw sensor data or showed the robot in a 2D overhead view in the context of its surrounding perceived objects. Applications like space exploration generated a strong impetus to develop more realistic virtual displays as early as 1991 [8]. These systems were tested terrestrially [6] and derivatives were ultimately used on the Mars pathfinder mission. Contemporary developments include more emphasis on sensor fusion [5] as well as efforts which display both forms of data in a less integrated but more useable way [15].

Our work presented in this paper is consistent with trends to provide more realistic views for the purpose of real-time control. However, our work emphasizes the construction of a novel sensing device which performs sensor fusion at the pixel level. This device has been designed to be well suited to solving the problem of virtualizing a real environment in real-time.

3 IMPLEMENTING PG SENSING

At some point in the future, flash lidar devices may be available which share apertures with color cameras in order to produce photogeometric data in hardware. Until that day comes, we find the value of PG data to be worth expending effort to produce it in whatever manner we can today.

Our implementation approach centers on the goal of producing an integrated data set of appearance and geometry data from two different sensors. The data may be organized arbitrarily but our two most common formats are camera-derived color data augmented with range, *rangified color* (RC), and lidar-derived range data augmented with color, which we call *colorized range* (CR) data.

Computational stereo vision is a natural RC modality because range is produced for every pixel in the reference appearance image. However, its utility in applications can be limited due to the relatively poor quality of the range data. This is often the case in our applications. Flash lidar sensors also continue to advance [1] but none yet meet our requirements for operation in outdoor environments. Conversely scanning lidar devices have been our preferred geometric sensing modality for two decades. Nevertheless, we will discuss PG sensing where the range data is provided by a scanning or a flash lidar.

In general, every appearance modality can potentially be paired with every geometry modality. Ideally, each sensor of a pair would image the same region of the scene as the other at the same resolution and frame rate from the same position. In practice, numerous technical issues arise due to the different attributes of the two sensors including:

- **Projective Geometry**. Lidar is often spherical polar, whereas cameras (and flash lidars) provide a perspective projection.
- **Resolution**. Scanning lidar typically produces 1% of the angular resolution (solid angle) of a camera so there can be up to 100 camera pixels for each lidar measurement.
- **Field of View**. Standard camera lenses, spherical mirrors, and lidar scanning mechanisms rarely provide the same field of view.
- **Location**. Displacement of one sensor center of projection or emission relative to another leads to parts of one view missing from the other – even if all other parameters match.
- **Frame Capture and Beam Scanning**. In cases where data is gathered on the move, each point of lidar data is captured from a different sensor position whereas all pixels in a camera frame come from a single position.

3.1 Establishing Pixel Correspondences

A basic property of cameras is their projective geometry which projects a 3D scene onto a 2D photosensitive sensor array. While the azimuth and elevation coordinates in the image are related to the equivalent directions in the scene, information about the depth of objects is lost when a camera image is formed.

Hence, the most valuable attribute of PG imagery is its recovery of the depth dimension which is lost when a real scene is imaged with a camera. This information is recovered by:

- establishing an association of lidar range points with camera pixels
- geometric transformations to convert lidar data to camera coordinates

For RC data, the color data is augmented with depth so that the result is an augmented image. For CR data the range data is colorized and the result is an augmented range image or point cloud. In either case, the mechanism to establish correspondences is the same. Consider Figure 2 which expresses the essence of the problem when both sensors are viewed from overhead.

For now, suppose that both sensors are stationary with respect to the scene and let us define a lidar “image” to mean the data produced by one sweep over the scene of the lidar scanning mechanism.

While it is not clear how to directly map color pixels onto a lidar data set, the reverse operation is conceptually straightforward. Hence both RC and CR datasets rely on a common procedure to establish correspondences. Let the letter L designate a coordinate frame attached to the lidar center of emission and let the letter C designate one at the camera center of projection. The homogeneous transform matrix that converts coordinates of a point from frame L to frame C is denoted T_L^C. Let the letter I designate row and column coordinates in the camera image plane. The projective transformation matrix that provides the image coordinates of a 3D point will be designated P_C^I. The homogeneous dimension will be omitted from vectors unless the matrices are written out. Under this notation, the camera image coordinates $\underline{r}^I = [x \quad y]^T$ of the point imaged by a lidar point $\underline{r}^L = [x \quad y \quad z]^T$ are:

$$\underline{r}^I = P_C^I T_L^C \underline{r}^L \qquad (1)$$

If the scene has sufficient 3D (non-planar) structure, the spatial separation of the sensors introduces characteristic problems of triangulation:

- **Missing parts**. Even with perfect field of view overlap, surfaces oriented perpendicularly (and invisibly) to the viewing direction of one camera may be visible to the other.
- **Depth ambiguity**. It is possible for the lidar to have ranged to a point on a background object that is behind a foreground object which was imaged by the camera.

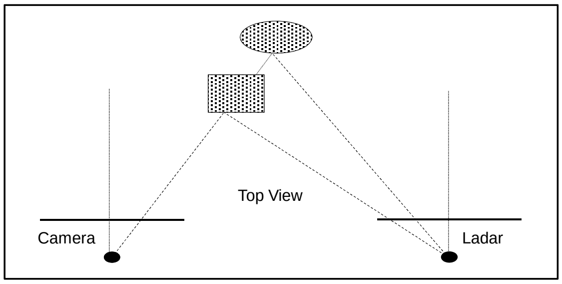


Figure 2: Multi Sensor Geometry and Depth Ambiguity. The camera measures the angle to objects whereas as the lidar measures angle and range. Due to the baseline separating the sensors, a lidar may image more than one object along the line of sight of a camera pixel.

While the first problem has no solution, the second can be solved by forming a depth buffer of all of the lidar data as viewed from the perspective of the camera image. All lidar data can be projected into bins which are sorted by depth or the processing may simply retain only the smallest range value in each bin. In either case, when two or more lidar pixels fall on the line through a given camera pixel, only the closest lidar point should be associated with the color pixel. All others are occluded and invisible to the camera so their color is unknown. While these triangulation issues cannot be eliminated entirely, they can be mitigated significantly by placing the two sensors very close together relative to the depths being imaged. However when the lidar is mounted on a moving vehicle, its continuous scanning process places limits on what can be achieved.

3.2 Forming Photogeometric Datasets

Given the correspondences between elements in each data set, either CR or RC data may be formed. The production of CR data using lidar is easiest to illustrate. In this case, the sensor intrinsic data format is a temporally ordered set of 3D points expressed in Cartesian or polar coordinates relative to the sensor center of emission. Each lidar point is simply augmented by the color of its associated camera pixel, if any. The color information might be the color of the closest camera pixel, the average over a region, or a block of pixels forming a small texture map.

In the case of RC data, the goal is to produce range data for every color pixel in a color image. Typical camera angular resolutions are 1 millirad whereas lidar is typically 10 millirad. Hence, once the lidar correspondences are computed, only 1% of the camera pixels can be expected to have associated lidar points. In other words, there will inevitably be holes in the coverage of the image by the range data. Small holes will be due to the reduced angular resolution of the lidar and larger ones due to missing parts or nonoverlapping fields of view.

When dense range data is desired, interpolation can be justified on the basis that the lidar is really providing the average range of the region of the scene that is spanned by a large number of camera pixels. The range data can be interpolated using the dilation operation of computer vision to fill small holes. The dilation radius can be related to the expected angular lidar footprint in the camera image. When both sensors are close together, the effect of surface orientation is minimal.

3.3 Sensor Configuration

Due to many considerations including the numerous robotic platforms that we construct annually and the desire to standardize solutions across programs, we have been continuously refining our photogeometric sensor concept for many years.

Two recent sensor designs are shown in Figure 3. For scanning lidars, we typically purchase an off the shelf scanning lidar which scans in one degree of freedom (called the *fast axis*) and then we actuate the housing in a second degree of freedom (called the *slow axis*) in order to produce a scanning pattern that spans a large angle in both azimuth and elevation. For flash lidars or stereo ranging systems, the interfaces to these devices are equal or similar to those of cameras so the process is more straightforward.

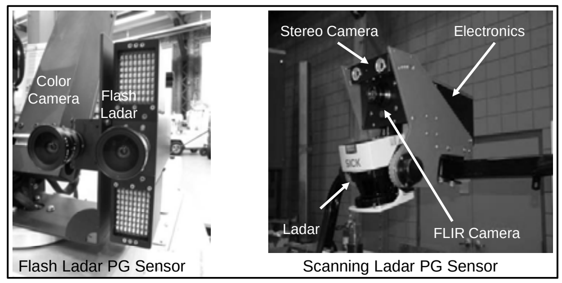


Figure 3: Two Custom Photogeometric Sensors. The device on the right fuses data from a commercial scanning lidar by SICK, stereo cameras, and a forward looking infrared (FLIR) camera. The device on the left fuses a PMD-Tec flash lidar with a color camera.

The interface to the composite device is a combination of fast Ethernet (used for high bandwidth data) and CAN Bus (used for low latency for control). One design goal is to render the composite device interface standard, high level, and easy to use.

The lidar pointing control system provides precisely timed feedback on the angle of rotation. This data stream is merged with the range and angle data coming from the lidar to form a 2D scanning lidar data stream. This stream is then optionally merged with any camera data and transmitted to the host computer system.

4 REMOTE MOBILE MANIPULATION

Mobile manipulation is a task for which human-robot interaction is often needed due to the difficulty of dexterous manipulation and the higher stakes associated with forceful interaction with the environment. While robots can often competently control their gross position, the final operations of the end-effector tooling may need to be performed with a human in the loop.

We recently conducted an effort to construct a mobile manipulation system that is analogous to commercial platforms and to endow it with a photogeometric sensor in order to study the benefits achievable when an operator designates a target to be manipulated on a video display [2]. Given the historical lack of range data, standard solutions to this problem include implementing a visual servo or using the robot navigation system to drive in the general direction designated until the operator issues a stop command.

However, in complex environments or in cases where the operator needs to direct attention elsewhere, it is more effective to have the robot decide when to stop. Furthermore, if autonomous obstacle avoidance is used, the robot can perform much more intelligently when it knows the precise 3D target for the manipulator end-effector. If manipulation and mobility are to be automatically coordinated, it again is necessary to know the 3D coordinates of the target. Hence, this is a case where it is valuable to have the human look at video while geometry (derived from the video) is communicated to the robot.

4.1 Platform Design

The robot used for these experiments was a modified LAGR mobile robot [9], fitted with a custom three degree-of-freedom manipulator arm and a gripper end-effector (Figure 4). The base vehicle has proved to be a very flexible research platform: in addition to over 40 standard models deployed at various universities, custom versions with LIDAR, metal detectors and omnidirectional cameras have been built.

The Photogeometric sensor consisted of two color video cameras, and the flash LIDAR unit of Figure 3 provided by PMDtec [14]. One camera is mounted on the manipulator arm near the wrist, for use during manipulation. The second camera and flash LIDAR unit is mounted to the shoulder yaw joint for driving and target acquisition.

The PMDtec sensor had a 30 Hz frame rate, a 64x48 range pixel array, a maximum range of 7.5 m, and was capable of operation in indoor and outdoor environments. The field of view was adjustable using standard C-mount optical lenses. For our experiments, a lens with a focal length of 4.2mm was used. This provided a 60 degrees horizontal and vertical field of view. By using a projective lens model, each pixel's range can be converted a 3D point in the workspace.

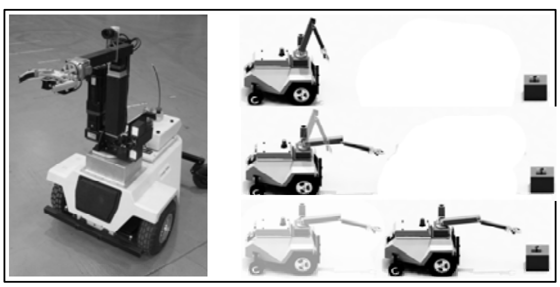


Figure 4: Test Platform for Click and Grasp. The LAGR robot platform was retrofitted with a manipulator and a custom photogeometric sensor.

4.2 Algorithm Design

We implemented a "Click and Grasp" function which allows an operator to click on a color image to designate a target, and have the system then a) recover the location in 3D space, b) navigate to within manipulation range, and then c) either grasp the target or put the end-effector as near as possible. In addition to our PG sensor, key aspects of the solution included control algorithms that coordinated the motion of the platform and the manipulator.

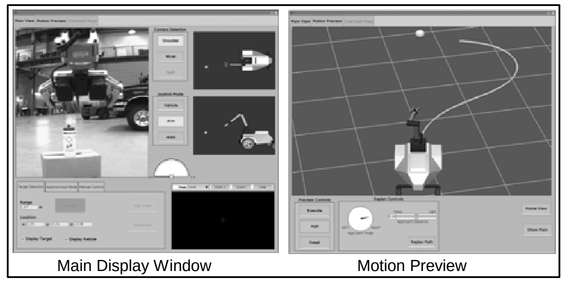


Figure 5: Operator Control Unit Display. The operator selects a target position in the left display. The system then plans and displays the entire motion and optionally asks for confirmation before execution.

4.3 Results

A custom graphical user interface was developed for the test (Figure 5). Two sets of experiments were performed to evaluate the effectiveness of the system in terms of increasing efficiency and usability.

In the first test, an object to be grasped was placed within reach of the manipulator and time to achieve the grasp was measured in various control modes.

Operator	Target 1 (secs. / errs)	Target 2 (secs. / errs)	Target 3 (secs. / errs)
Auto	28.27 / 0	22.52 / 1	24.20 / 0
Expert (JS)	41.92 / 0	40.84 / 0	45.73 / 0
Expert (WS)	47.73 / 0	32.75 / 0	45.37 / 0
Novice (JS)	38.43 / 0	36.95 / 0	32.70 / 0
Novice (WS)	29.21 / 1	22.85 / 0	33.36 / 0

Table 1: Pick Up of Object Within Reach of Manipulator. The operators were allowed to operate the manipulator in both joint-space (JS) and end-effector workspace (WS). "Auto" corresponds to using the automatic "click and grasp" system.

The automatic system was able to accomplish the task significantly faster than both (expert and novice) operators using manual teleoperation, and workspace controls significantly improved the completion time of the novice operator. On average, the automatic click and grab system was able to perform the static manipulation task 13% faster than both users.

Operator errors were noted whenever the test target was knocked over by the manipulator. The test was then reset and the operator permitted to retry. If a failure was due to insufficient grasping force at the end effector, it was not counted as an operator error. Our intent was to focus on characterizing the utility of the system as an aid to the precise positioning of the end effector.

In the second test, the object was outside the manipulator workspace so the platform had to be moved in either an automated or manual fashion.

Operator	Target 1 (secs./errs)	Target 2 (secs./errs)	Target 3 (secs.)	Target 4 (secs.)
Auto	37.08 / 0	68.43 / 2	54.61 / 0	56.21 / 1
Expert	48.37 / 0	40.77 / 0	54.75 / 0	43.00 / 0
Novice 1	45.42 / 0	48.98 / 0	40.52 / 0	32.42 / 0
Novice 2	49.52 / 1	49.43 / 0	43.17 / 0	35.48 / 0

Table 2: Pick Up of Object Outside Reach of Manipulator. The automatic "Click and Grab" system performed comparably to the human operators.

Work-space controls reduced both the time required to complete the task as well as the number of errors made. On average across four trials, operators reduced their number of errors from three to one, and reduced their completion time by 11%. The results demonstrate that the autonomy and perceptive capabilities of our system eases the workload on the operator while increasing task efficiency.

Results from the automatic system were potentially limited by the accuracy of the flash lidar range and co-registration. At short ranges, accuracy was sufficient to reliably grasp an object. However, at longer ranges, errors were large enough to cause manipulation errors. Instead, "click and grab" at long range required several operator interventions to re-designate the target once the base had positioned itself within range.

5 MOBILE ROBOT TELEOPERATION

Effective operation of any mobile platform without direct line-of-sight is intrinsically difficult to achieve. In video-based teleoperation, the loss of peripheral vision caused by viewing the world through the "soda straw" of a video camera reduces driving performance and increases the operator's frustration and workload. Wireless communication links are also subject to dropouts and high levels of latency. Their bandwidth limitations typically cause a large reduction in image quality relative to the fidelity of the underlying video cameras.

When the robot undergoes significant or abrupt attitude changes, the operator response may range from disorientation, to induced nausea, to dangerous mistakes. In contexts where the operator is also in danger, the need for high attention levels deprives operators of the capacity to pay attention to their surroundings. Wireless communications issues and difficulty controlling the robot also increases time on task and increases the time required to become a skilled operator.

5.1 PG Sensing for Autonomy

We have been working on improved operator displays for at least a decade on our robot autonomy programs [3]. PG sensing was originally motivated by its capacity to disambiguate natural obstacles and non-obstacles of the same shape (such as a rock and a bush) by examining their color signatures (see Figure 6). Once the data was available for use in autonomy however, we began to produce specialized point cloud displays and quickly recognized the potential of the PG data for human interfaces.

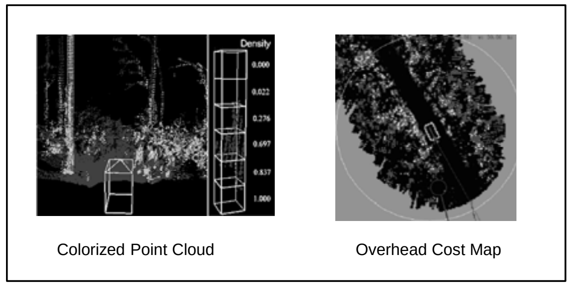


Figure 6: Original Engineering Displays of PG Data on Autonomy Programs. The display of traversability / cost or elevation from an overhead display (right) is traditional in robotics. In recent years, colorized point clouds have also been used. The evolution of the left figure toward photorealism was a natural extension of ongoing efforts.

5.2 3D Video

Photogeometric sensing enables a new capacity to address many of the problems described in the introduction to this section by providing a photorealistic, synthetic, line of sight view to the robot based on the content of geometry-augmented real-time video feeds. The operator experience is equivalent to following the robot in a virtual helicopter that provides arbitrary viewpoints including an overhead viewpoint and the over-the-shoulder view that is popular in video games.

The fusion of video and geometry produces a database whose content is much closer to a computer graphics rendering database than basic video. If the geometry is converted to faceted surfaces and the imagery is converted to textures, the PG sensor data has been converted to a rendering database. If this conversion is performed in real-time, a kind of hybrid *3D Video* is produced which can be viewed from arbitrary perspectives while exhibiting the photorealism and dynamics of live video.

If the sensors are omnidirectional and/or if the system remembers and integrates the rendering primitives over time, the net result is the real-time virtualization of the scene which enables the operation of the robot quite literally as if it were a 3D video game.

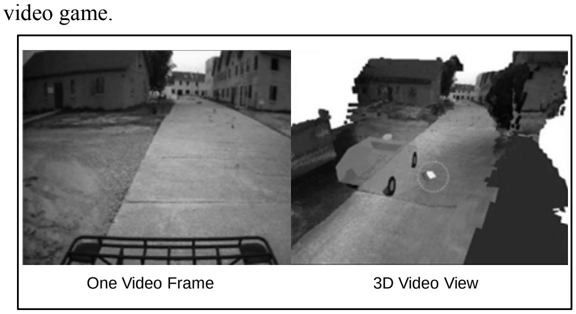


Figure 7: 3D Video View of a Mobile Robot. Left: A video frame produced from a camera on a moving vehicle. **Right**: The 3D Video view produced from all of the video that has been received in the last few seconds by the vehicle. The operator can

look at this database from any angle, at any zoom, while it continues to be updated in real time.

The ability to synthetically generate a viewpoint via computer graphics leads to the following capabilities:

- A natural mechanism to introduce virtual operator aids into the display.
- The capacity to zoom into objects of interest and view them from different perspectives (for example, from above or from the side) from just a few inches away.
- The capacity to have multiple operators cooperate from multiple views, perhaps even using cooperating robots.

5.3 Results

The goal of 3D Video technology is to increase an operator's situational awareness of the vehicle being controlled, thereby reducing operator errors and increasing the speed with which tasks are completed.

We conducted an operator performance assessment over a period of one week involving five operators of different skill levels. The participants averaged 20 years of automobile driving experience. Three subjects had prior experience teleoperating a live vehicle, including one with a 3D Video system. Two of these subjects had participated in one other experiment, while the other had extensive experience, teleoperating a vehicle in many experiments. Three subjects had minimal experience teleoperating a *simulated* vehicle (two of these included in the group with live vehicle experience). Four subjects had been playing driving-based video games for an average of 13 years, with one subject playing as often as a few times per week. One subject had never played a driving based video game.

The test platform was a John Deere eGator vehicle retrofitted for remote control and teleoperation. Participants completed four test conditions, which were counter-balanced across participants to minimize order effects related to course and Operator Control Station (OCS) familiarity:

1. Manually drive from seat
2. Basic Teleoperation with live video
3. Teleoperate with 3D Video - without motion prediction
4. Teleoperate with 3D Video - with motion prediction

Motion prediction refers to a method used to alleviate the effects of video latency. We use the most recent navigation state received from the robot and predict the robot position based on the terrain shape and the history of operator inputs. Due to latency in sending commands to the robot, the instant of time being predicted is not "now" but rather the moment in the future when the commands are predicted to arrive at the vehicle. In principle the display will then respond instantly to operator inputs and it will correspond to a point in time slightly ahead of where the vehicle is now. The availability of lidar data makes it possible to predict robot motion to relatively high accuracy compared to the alternative of ignoring the latency.

The course consisted of a paved roadway with traffic cones set up to guide drivers at particularly ambiguous areas such as intersections. Course features included slaloms, decision gates, and discrete obstacles as a series of loose and tight turns.

Performance metrics included course completion time, course accuracy, average speed and errors as well as subjective input on workload [7], impressions of the system and recommendations for future improvement. Errors were defined as hitting a cone, (having the vehicle emergency-stopped before) hitting an obstacle along the edges of the course (concrete barriers, fences, and hay bales occurred sporadically along the perimeter of the course), or deviating from the defined region of the course (driving off the road). In the end, course accuracy was not measurable due to data collection equipment availability.

Course Completion Time Results: As the figure below indicates, 3D Video enabled operators to complete the course faster than basic teleoperation: 3D Video alone led to completion times approximately 20% lower, while times were 30% lower when 3D Video was combined with motion prediction (MP). As expected, manual driving (in the vehicle) is still far superior, with course completion time approximately 75% lower than basic teleoperation.

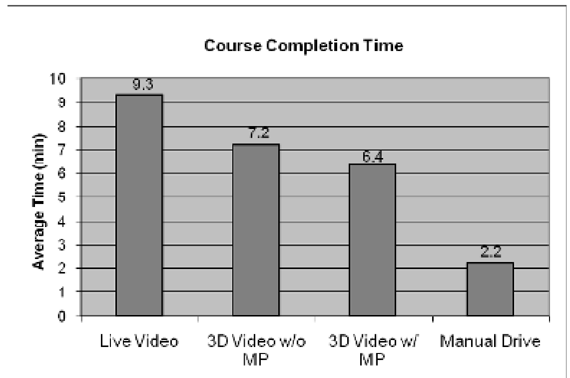


Figure 8: Average Course Completion Times

Speed Results: The benefit of 3D Video follows the same trend as completion time. Basic teleoperation achieved 1.0 m/s average speed, while 3D Video alone led to 30% faster driving, and 3D Video with motion prediction increased speed by 50%. Manual driving was more than three times faster.

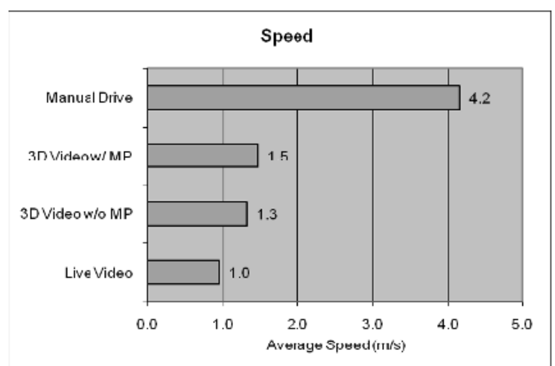


Figure 9: Average Speed

Number of Stops: One interesting repetitive event was operators choosing to stop the vehicle. This was a common response when relevant information was not available due to limited field of view or because latency disoriented the operator. 3D Video configurations reduced the frequency of stopping by 43% when compared to basic teleoperation. No drivers stopped during the manual driving configuration.

Error Rate: Fewer errors were made with the 3D Video than basic teleoperation. With 3D Video alone, the error rate dropped by almost 50%, while the error rated dropped by about 20%

when 3D Video was combined with motion prediction. Manual driving is again the gold standard, with an error rate approximately 75% lower than basic teleoperation. Interestingly, the course was sufficiently complex that drivers did commit errors even with manual driving. The average rate was 2.4 errors per run, and every driver committed at least one error over the course.

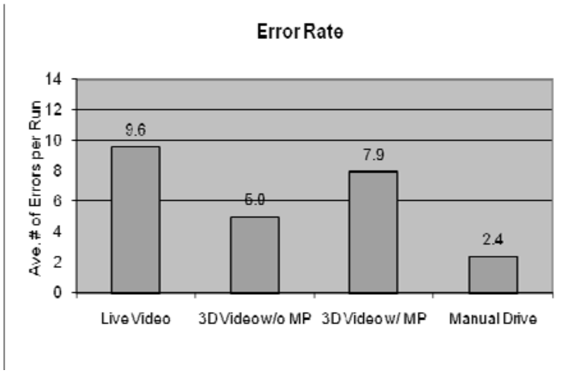


Figure 10: Error Rate

One result is particularly interesting: drivers made markedly fewer errors without motion prediction than with. In other metrics, motion prediction generally shows a small but positive benefit, while here, the contrast is substantial. Analysis of the motion prediction system likely explains why: a combination of i) variable latency invalidating the constant-latency model used in the software, ii) sub-optimal vehicle model parameters and iii) inaccuracies in the pose data, all contributed to errors in motion prediction that were at times substantial (relative to the tolerance of many of the course decision gates, for example).

Workload: The NASA TLX workload questionnaire was administered after each run, allowing operators to rate perceived mental demand, physical demand, temporal demand, own performance, effort and frustration associated with each driving condition. Overall workload scores indicate the least amount of workload was required with the 3D visualization system alone. As expected, the highest workload was achieved with live video, while 3D Video with motion prediction and manual drive were rated similarly. In general, manual driving workload was rated higher than expected. This may be due to the physical effort required to use the eGator steering wheel and a lower than anticipated perceived performance rating.

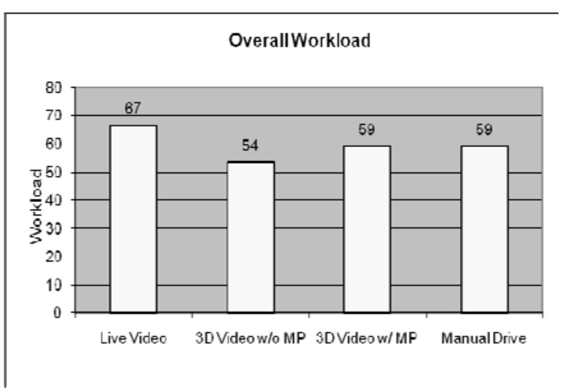


Figure 11: Overall Workload

Looking at the dimension scores associated with overall workload, differences between driving conditions become more apparent. Live video required significantly more mental demand than other driving conditions, as well as higher temporal demand, perceived effort and frustration levels. Temporal demand ratings were very close, which is not surprising given drivers were told to complete the course as quickly as possible, thereby creating time-based workload across all conditions. 3D visualization conditions were rated similarly, but frustration levels were higher without motion prediction. Drivers reported the lowest physical demand with the 3D Video conditions.

Exit Interview: The exit interview was completed with each participant at the conclusion of all runs. The most commonly requested improvements for basic teleoperation include decreased latency, higher video frame rate and more cameras or unique viewpoints. Participants also mentioned better resolution, wider field of view, and an indication of vehicle position in the video frame, which would allow them to drive through tight spaces. In general, operators wanted the ability to judge where the vehicle is positioned in the world by having a direct reference to all objects in the environment.

Participants felt the greatest strength of the 3D Video is the vehicle model presented within the video. The model made it easier to recover from mistakes and allowed operators to judge upcoming course events with respect to the vehicle, thereby allowing them to respond to the environment more accurately. “I could go faster between events and then slow down before an event. I could time the slow down better.” 3D Video also provided a wider field of view, latency compensation, and selectable viewpoints. These features provided a “less stressful” environment and reduced the amount of time spent “paying attention to the vehicle,” potentially freeing up time for other vehicle control and mission-related tasks.

The following artifacts were present in 3D Video: “video jittering around corners, straight lines in the middle of a road bending, cones disintegrating and appearing flat on the surface, 3D objects smearing as vehicle drove by, and square pixels appearing at the edge of imagery.”

3D Video improvement suggestions include reducing artifacts, a higher video frame rate, improvements in latency compensation, and a wider field of view for turns. A higher frame rate was suggested to make driving at a higher velocity easier.

The final portion of the exit interview allowed participants to rank their preferences for driving condition and 3D Video viewpoints. Manual driving was preferred, followed by 3D Video with motion prediction, 3D Video without, and Live video. Three viewpoints were available within the 3D Video: native camera, over-the-shoulder, and overhead (bird’s eye view). The overall preference for viewpoints was unanimous: over-the-shoulder, followed by Overhead (Bird’s Eye View) and native.

Comments indicate bird’s eye view was useful when navigating left or right for a short distance, such as in a slalom, and native location was useful if driving on straight roads for a long distance. Over the shoulder was more or less the “all purpose” preferred viewpoint.

Summary: In perhaps the most significant metric of task completion time, 3D Video showed improvements of approximately 20-30% compared to standard teleoperation.

Other metrics showed improvement as well, with average speed increasing 30-50 % and error rates dropping by 20-50%.

6 CONCLUSIONS & FUTURE WORK

As long as humans use displays of data generated on a remote device in order to control it, the sensors deployed on the device will be used to produce those displays. While single use sensors are common, the dual use of robot perception sensors for autonomy and visualization is already well established.

This paper has proposed a method to expend significant engineering effort in order to produce a virtual sensor with the ideal characteristic of reducing, in real-time, the environment around a remote device to the essence of a computer graphics rendering database. In other words, photogeometric sensing has the capacity to virtualize reality to produce displays with both the photorealism of video and the interactivity of a video game.

PG sensor technology can be transformative for certain human-machine interface tasks, providing solutions to problems for which there has been little hope of significant progress for some time. We have produced many instances of PG sensors in recent years and deployed them in diverse applications. In the two discussed here, user studies have verified substantial gains in the effectiveness of the man-machine system.

PG sensor technology introduces entirely new and highly effective approaches to latency compensation and video compression which have not been elaborated here for reasons of space and focus.

We are reasonably convinced that PG sensing is a sensor modality of choice with unique advantages that enable a new level of shared mental model between a robot and a human. Based on it, communications between the two can become less frequent, more terse, and more precise.

We have produced PG sensing by integrating distinct geometry (lidar) and appearance sensors (camera) into a virtual unit. Over time, the eventual development of high accuracy shared aperture sensors which are integrated in hardware at the pixel level seems inevitable.

7 ACKNOWLEDGEMENTS

The colorized ranging sensor concept described here was based on a concept originally developed under funding from John Deere Corporation. The work in mobile manipulation was funded by the US Navy EOD Technology Division (NavEOD Tech Div). The work in applying 3D Video to robot teleoperation was funded by the US Army Tank Automotive Research, Development, and Engineering Command (TARDEC). The autonomy work described in Figure 6 was funded by DARPA.

REFERENCES

[1] D. Anderson, H. Herman, A. Kelly, "Experimental Characterization of Commercial Flash Lidar Devices", In Proceedings of International Conference on Sensing Technologies, November 21-23, 2005 Palmerston North, New Zealand.

[2] D. Anderson, T. Howard, D. Apfelbaum, H. Herman and A. Kelly, "Coordinated Control and Range Imaging for Mobile Manipulation" 2008 International Symposium on Experimental Robotics.

[3] R. Chatila, S. Lacroix, T. Simion, M Herrb, "Planetary exploration by a mobile robot: mission teleprogramming and autonomous navigation" Autonomous Robots, 1995.

[4] C. Dima, N. Vandapel, and M. Hebert, "Classifier Fusion for Outdoor Obstacle Detection", International Conference on Robotics and Automation, April, 2004, pp. 665 - 671.

[5] T. Fong, C. Thorpe, and C. Baur, "Advanced Interfaces for Vehicle Teleoperation: Collaborative Control, Sensor Fusion Displays, and Remote Driving Tools. Autonomous Robots 11, pp 77--85, 2001.

[6] T. Fong, H. Pangels, D. Wettergreen , "Operator Interfaces and Network-Based Participation for Dante II", SAE 25th International Conference on Environmental Systems, San Diego, CA, July 1995.

[7] S. G. Hart and L. E. Staveland (1987). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock and N. Meshkati (Eds.), Human mental workload. Amsterdam: Elsevier.

[8] B. Hine, et. al., "The Application of Telepresence and Virtual Reality to Subsea Exploration", the 2nd Workshop on Mobile Robots for Subsea Environments, Proceedings ROV '94, Monterey, CA, May 1994.

[9] Jackel, L., Krotkov, E., Perschbacher, M., Pippine, J., Sullivan, C.: The DARPA LAGR program: Goals, challenges, methodology, and phase I results. Journal of Field Robotics 23, 945–973 (2006)

[10] A. Kelly et al., "Toward Reliable Off-Road Autonomous Vehicles Operating in Challenging Environments", The International Journal of Robotics Research, Vol. 25, No. 5–6, pp. 449-483, June 2006.

[11] R.A. Lewis and A.R. Johnston, "A Scanning Laser Rangefinder for A Robotic Vehicle" Proceedings of I.ICAI~5, 1977.

[12] J. Ryde, H. Hu, "3D Laser Range Scanner with Hemispherical Field of View for Robot Navigation", Proceedings of the 2008 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, July 2 - 5, 2008, Xi'an, China.

[13] D. A. Messuri and C. A. Klein, "Automatic Body Regulation for Maintaining Stability of a Legged Vehicle during Terrain Locomotion", IEEE Journal of Robotics and Automation, RA-1, pp 132-141, Sept 1985

[14] T. Möller, H. Kraft, J. Frey, M. Albrech, R. Lange.: Robust 3D measurement with pmd sensors. In: Proceedings of the First Range Imaging Research Dat at ETH Zurich (2005). ISBN 3-906467-57-0

[15] B. Ricks, and C. Nielsen and M. Goodrich. "Ecological Displays for Robot Interaction: A New Perspective". Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004.

[16] T. Sheridan, "Human Supervisory Control of Robot Systems", Proceedings of the 1986 IEEE International Conference on Robotics and Automation, Apr 1986.

[17] G. Terrien, T. Fong, C. Thorpe, and C. Baur. "Remote driving with a multisensor user interface". In Proceedings of the SAE ICES, Toulouse, France. 2000.

A WOz Framework for Exploring Miscommunication in HRI

Theodora Koulouri[1] **and Stasha Lauria**[1]

Abstract. The aim of this paper is to investigate management of miscommunication in spontaneous interaction between a human and a speech-enabled robot. The paper describes initial results of an exploratory WOz study, in which pairs of naive participants interacted and collaborated in a navigation task. The study is motivated by how humans achieve mutual understanding as well as previous research conducted in the area of spoken dialogue systems. The dialogue and error handling capacity of the wizard is incrementally impaired towards the capabilities of a system in three experimental conditions. Preliminary analysis of the data reveals the necessity to endow the robot with richer error management resources and point to the efficacy of less explicit error handling strategies. Analysis of the data and further experimentation are currently performed and expected soon to shed light to some of the intricacies and unique characteristics of HRI.

1 INTRODUCTION

1.1 Personal Robotics

Recent progress in robotic technologies brings closer the vision of ubiquitous and commercial use of robotic systems. This poses new challenges for the vibrant and young field of Human Robot Interaction (HRI). Personal robots possess the highest expected rate of growth [1] and as non-expert user groups are targetted, more natural and rich interfaces are on demand. Thus, there are numerous prototypes equipped with natural language interfaces (NLI) and promising results have been generally reported [1]. Nevertheless, there remain many challenges which pertain to the development of a NLI between a human and a robot.

1.2 Natural (mis)Communication in HRI

Natural language is infinitely novel and ambiguous. Moreover, language in use, i.e., language that emerges in dialogic settings, is never perfect. In fact, error-free communication is now held to be an ideal rather what happens in everyday human interaction. Yet, human interlocutors manage miscommunication so efficiently that it rarely becomes an explicit focus in the dialogue. Additionally, the performance of state-of-the-art NLP technologies still leaves much to be desired. As a consequence, problems in understanding, uncertainty and out-of-grammar words will always occur in human-machine interaction. Most current spoken dialogue systems (SDS) produce prompts such as "Please repeat" and "I do not understand" to signal and resolve problems. Such strategies are, nevertheless, insufficient to handle all kinds of miscommunication, which has ultimately a great impact on performance and user experience. This issue has first been identified in research in the area of spoken dialogue systems, which has mostly dealt with information-seeking, telephone-based applications. But it is exacerbated in the area of personal robotics in which users are naive not only about the linguistic but also the functional competence of the robot [2]. It is evident that errors in understanding and execution give rise to many safety concerns which is not usually the case in other domains. Moreover, it has been argued [1] that language-endowed robots face higher expectations by people, who attribute human-like linguistic capabilities and intelligence to them. Thus, as the occurrence, scope and forms of miscommunication increase, the need for an approach that enriches the robot with a greater repertoire for dealing with problematic understanding also increases.

The aim of our research is to develop a natural framework for handling miscommunication in HRI. As people are extremely apt in preventing and repairing problematic understanding, the approach in this study is to explore and build on the principles of human error handling. This paper describes the first steps in identifying consistent linguistic behaviour when human users perceive communication failure within the context of HRI.

2 BACKGROUND AND MOTIVATION

2.1 Past Work

The current work extends previous research by the universities of Plymouth and Edinburgh [3]. The project explored Instruction Based Learning (IBL) and aimed to enable naïve human users to instruct a personal robot to perform a navigation task by means of natural language. The robot is equipped with a built-in knowledge of some basic navigation actions. On encountering a novel route description, the robot engages in a dialogue with the user who explains and decomposes the novel route into known actions. Subsequently, the route is incorporated to the robot's knowledge base for future use. This enables the robot to learn and execute increasingly complex tasks. This work demonstrated that IBL is a viable architecture for developing personal robots with capabilities of learning. However, evaluation of the system revealed the complexity of adding even simple interactive mechanisms to the robot [4].

2.2 Models of Communication

[1]Dept. of Information Systems and Computing, Univ. of Brunel, UB8 3PH, UK. Email: theodora.koulouri@brunel.ac.uk

Empirical studies of human-human interaction (HHI) describe communication as a collaborative process in which participants are continuously establishing that what has been said has also been understood. During interaction, interlocutors can be in any of these states [5]:

Table 1. States of Understanding. B, A and u stand for hearer, speaker and utterance, respectively.

State 0	B didn't notice that A uttered any u
State 1	B noticed that A uttered some u (but wasn't in state 2)
State 2	B correctly heard u (but wasn't in state 3)
State 3	B understood what A meant by u

Communication is achieved if interlocutors are in state 3. According to the model, listeners choose repair initiations that assert their current state of understanding. For example, "What?" is an assertion of being in state 1 and "Which building?", a presupposition of state 2.

Along the same lines, [6] and [7] propose a four-level model of communication which has been unified by [8] and is shown below. Miscommunication can occur in any of these levels:

Table 2. Levels of Communication (adapted from [8]).

Level 1	Securing Attention
Level 2	Utterance Recognition
Level 3	Meaning Recognition
Level 4	Action Recognition

2.3 Research in Miscommunication

2.3.1 Sources of miscommunication

Miscommunication covers three categories of problems in interaction; misunderstandings, non-understandings and misconceptions [9]. Misunderstandings occur when the hearer obtains an interpretation which is not aligned to what the speaker intended him/her to obtain and may not be readily detected. This category generally leads to direct corrections by the speaker. Non-understandings occur when the hearer obtains no interpretation at all or too many. This category could also include cases in which the hearer is uncertain about the interpretation which he/she obtained. Non-understandings are of special interest to our study as this type of problem triggers repair initiations (e.g., clarification requests) by the hearer. Last, misconceptions occur when the interlocutors' beliefs of the world clash.

2.3.2 Clarification requests

Clarification requests can be broadly defined as the dialog acts employed by the hearer to signal a problem in understanding. Most taxonomies of clarification requests [10, 11, 12] are motivated by the models of HHI discussed in Section 2.2. Namely, it is maintained that the choice of clarification request indicates the highest level of understanding currently available to the listener and signals the information required for understanding to eventually occur. Thus, there is the expectation that speakers modify their original utterance as a response to a clarification request.

Furthermore, "generic" repair utterances such as "What?", "Pardon?" and "Please repeat" (that is, the default error handling strategies of a SDS) are less informative as regards to the source of the problem compared to reprises (such as "Go where?"). According to [12], the latter type of clarification requests accepts a part of the problematic element and is less severe for the dialogue. A study in HHI [8] compared "What's" and reprise fragments and found that the former caused more dialogue disruption and impaired the coordination between participants.

2.3.3 Error handling strategies in SDS

In the area of SDS, there have been considerable efforts to build systems with more intelligent handling of errors. In [13], the problematic utterance was classified in one of twelve classes of errors and targetted help was provided by the system. In [14], the system gave feedback with the recognition hypothesis, a diagnosis of the problem and similar in-coverage examples of what to say. Both studies report more successful interactions. In [15], an extensive research in error recovery strategies from non-understandings is presented. They defined a set of error handling strategies and discovered a relation between them and recovery from errors. They provided evidence for benefit in performance by incorporating a smart policy of selecting strategies.

In previous studies [16, 17, 18], variations of Wizard of Oz (WOz) simulations were performed in order to discover how humans handle automatic speech recognition (ASR) errors. Zollo [16] lists a number of negative and positive feedback strategies such as reprise sluices and fragments, confirmation requests, simple acknowledgements etc. Moreover, Skantze [17] found that explicitly signalling non-understanding is not the most common strategy used by humans to handle errors, but providing task-related information results in more successful interactions. Findings from [15] and [18] for different domains seem to resonate with these conclusions. The aforementioned studies suggest that dialogue systems that make miscommunication another contribution circle and not a focus of the dialogue will enjoy higher success rates in terms of performance and user experience.

There have been several successful implementations of prototype SDS that incorporate clarification requests [11, 12] and more sophisticated error handling. However, it remains an open question how these features would operate in a robot which is perceptually grounded to its environment.

2.4 Challenges in HRI

Research in error handling focuses on uncertainties arising from poor speech recognition, which is justified by the fact that it is the source of the large majority of errors. However, a significant number of errors are attributed to out-of-grammar expressions and requests beyond the functionality of the system [15].

In the interaction with a personal robot acting in the same dynamic environment as its user, one should expect a different distribution of errors and possibly new sources of errors. To the knowledge of the authors there has not been an analysis of sources of errors in speech-enabled robots and how each of these affects the global performance of the robot.

n this section the unique characteristics of HRI that make NLI an even more challenging endeavour are outlined. First, the domain of interaction is wider than of SDS. Moreover, as HRI is a new field, users are more naive about what the robot can understand and do. It has also been suggested that they would expect human-like abilities [1]. As a consequence, misunderstandings and non-understandings are more likely to occur. Expanding the grammar by collecting larger corpora is a difficult and resource-demanding task with disproportionate outcome. Therefore, the robot should be capable of informing users of its competences in a natural manner [2]. In the face of problematic understanding, the robot should be able to act in a way that ensures the safety of the user but also the smoothness of the interaction.

On the other hand, models of HHI predict that co-presence will increase the perception of mutual knowledge [19] and users are more likely to refer to objects and entities in a way that a robot without human-like vision capacities is unable to resolve. Moreover, real physical environments are dynamic, even more so when the robot is mobile. Therefore, referential resolution obtains an elevated status in the dialogue manager of the robot, which is now required to be able to communicate and negotiate changes promptly and effectively.

2.5 Methods of Data Collection

It could be argued that since natural interaction is the end, data from human dialogues should be the starting point. However, it is well established among linguists that speakers adapt their linguistic behaviour according to the perceived characteristics of the hearer. For instance, we do not talk to a five-year old child in the same way we talk to a colleague. This might explain the differences in interaction patterns observed in HHI and HCI [20]. Moreover, communication in conversational settings comes with an abundance of shared assumptions and knowledge which are neither transparent nor relevant to the researcher. Another solution could be to use data from interactions with current spoken dialogue systems. This, however, seems to be of little utility as our aim is to build the systems of the future.

In order to collect data from a variety of interactions and phenomena as well as test functionality not yet implemented, NLI developers set up WOz experiments. In these experiments, a human operator ("the wizard") emulates the system (or parts of it) and interacts with a user who is under the impression that he/she is talking to a machine [20].

3 METHOD

This study aims to identify the strategies that humans use when their communicative ability and the information available to them are restricted in a way similar to artificial systems. To serve our purposes, we devised a series of WOz experiments which deviates in certain respects from the typical WOz methodology. Previous research that employed this method [15, 16, 17, 18] served as the motivation and basis for our design. WOz simulations have also been conducted in the area of HRI [3, 21]. However, to the authors' knowledge, none of them explored miscommunication and error handling.

In a WOz experiment the wizard is the trained experimenter. However, this seems to introduce a bias as the experimenter controls the interaction. As the object of this study is the wizard's dialogue actions, the experiments involved *two* naive subjects, that is to say, both the user and wizard are naive. A justification of this choice comes from [17] who maintains that "this experimental setting lacks the control that the consistent behaviour of a trained operator would give. Still, this method may be good for explorative studies, which aim at finding new ideas on dialogue behaviour, and especially on how error situations could be handled." The study explored recovery strategies from ASR errors. However, both participants were fully informed, hence, there was no wizard involved.

The experimental design described here is also largely motivated by the WOz paradigm for dialogue systems proposed by Levin and Passonneau [22] which draws on two methods used in AI, namely, ablation and comparison. In their design, the communicative resources of the wizard are incrementally restricted. The different conditions are compared in order to isolate the properties of the dialogue system that most affect the overall performance. Their paper also exemplifies the application of the method for exploring error handling strategies.

Our study involves three experimental conditions for data collection. In particular,

- The wizard simulates a super-intelligent robot and interacts with the user using unconstrained natural language (henceforth, referred to as Condition 1).
- The wizard selects from a list of utterances that point to the source of the problem (in the utterance, meaning or action level, see Section 2.2) but can also type in a clarification question or provide task-related information (Condition 2).
- The wizard is fully restricted to use the same limited set of utterances as a typical dialogue system. In case of problematic understanding, the wizard signals that there is some kind of problem (Condition 3).

3.1 Domain of the Experiment

The domain of the experiment is navigation on a miniature town which is similar to the Map Task [23]. As Brown [23] points out, in the Map Task, each subject has two overt sources of information: what the other speaker says and what is on his/her map. Thus, the participants are given the opportunity to interact with each other in a relatively natural manner, while controlling the information available to them at any given point in the dialogue. The analyst has access to the records of what each participant does and says, which allows for a degree of understanding of how people are tackling the task or the problems that arise, and where these arise. Yet, even in task-oriented interaction, repair instances are not frequent.

In a small-scale study, like the one described in this paper, the scarcity of such data would have prevented us from making any reliable inferences. Thus, the interface between the participants had to be further degraded (as explained below). In our study, the user guided the robot to six destinations. The user had full access to the map whereas the wizard could only see a small fraction of the map of the area surrounding the robot, so he/she had to rely on the user's instructions on how to go to a location. In addition, the user could not see the robot itself, but only its surroundings. Therefore, both participants needed to collaborate and exchange information in order to complete the task. In this sense, the task falls under the category of problem-solving tasks, as opposed to information-seeking tasks commonly employed in the development of SDS. It is interesting to discover to what extent knowledge gained from error handling in the domain of information-seeking dialogue systems can be applied to a different domain of interaction with a robot.

3.2 Set-up

A custom java–based system[2] was developed for the experiments consisting of the map and the messaging box. The wizard's interface was modified according to each experimental condition but the user's interface remained the same. The two applications were linked together using the TCP/IP protocol, sending and receiving coordinates and messages over a LAN. However, there are no real constraints of location as the computers can also be connected via the Internet. Moreover, the system kept a log of the interaction but also, for every message exchanged, the coordinates of the robot at that given moment were recorded. This enabled us to monitor the wizard's level of understanding by comparing each instruction with how it was actually executed.

For all conditions, the wizard's map contained only a small fraction of the map with the area that surrounded the current position of the robot. All buildings were shown as yellow squares. The robot was displayed as a red circle with a yellow "face" and was operated by the wizard by pressing the arrow keys on the keyboard. The dialogue box was on the lower part of the screen. The messages from the wizard were shown on the upper part of the box (in green) and the user's messages on the bottom as well as a history of the user's previous messages. As explained in 2.1 above, the actual robot has the ability to learn and remember previous routes. To simulate this ability, once a route was completed, in the next task, a new button appeared on the right side of the wizard's screen that represented the newly learnt route. When the user requested to take a known route, the wizard clicked on the corresponding button and the robot automatically executed the route.

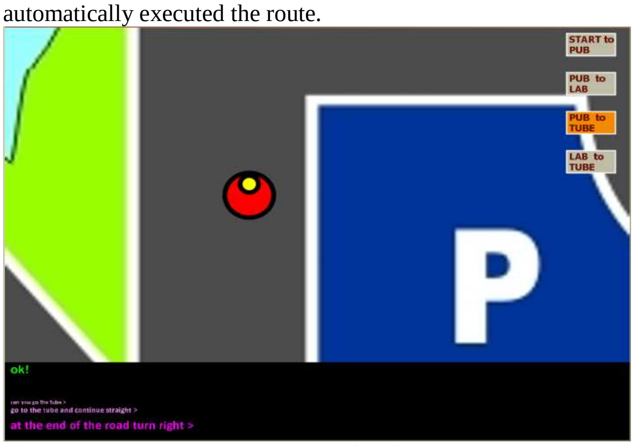


Figure 1. The Wizard's Interface for Condition 1.

In the interface version for Condition 1, the wizard could freely type messages and send them to the user like in a typical messaging application. Figure 1 displays a screenshot of the interface. Note that that the robot has already completed and "learnt" four routes.

In the version for Condition 3 (Figure 2), the wizards were deprived of typing their own messages. There was a set of buttons on their messaging box which the wizard clicked and the corresponding canned response was sent to the user. These responses were "hello", "goodbye" and "ok". In addition, there was a button denoted with a "?". The wizard was instructed to use this button to inform the user that there was some kind of problem in understanding or executing. This button randomly generated any of the following responses: "Can you please repeat that?", "Sorry, I don't understand", "What?" as well as a response that contained a fragment of the user's previous message (e.g., "Turn where?"). The wizard had no control over or any a priori knowledge of which response would be sent.

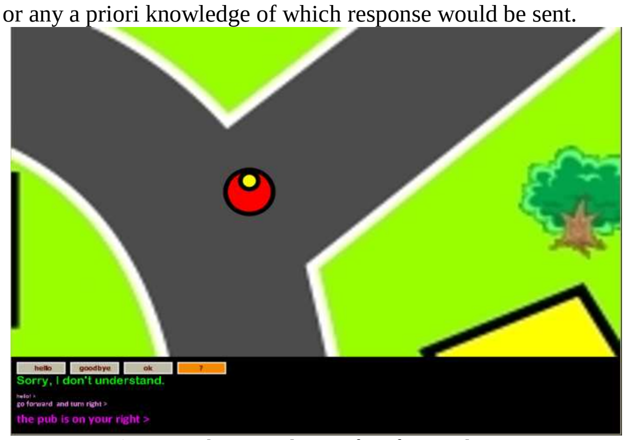


Figure 2. The Wizard's Interface for Condition 3

The interface version for Condition 2 (Figure 3) allowed for two ways of interaction with the user. First, the canned response-based interaction; in particular, the wizard could click on "Hello", "Goodbye", "Yes", "No", "Ok" and the problem-signalling buttons "What?", "I don't understand" and "I cannot do that"[3] to automatically send these messages to the user. The second way was to click on the "Robot Asks Question" and "Robot Gives Info" buttons so that the wizard could type in his/her own messages; the wizard was instructed to use the former to request clarification and the latter to provide the user with information.

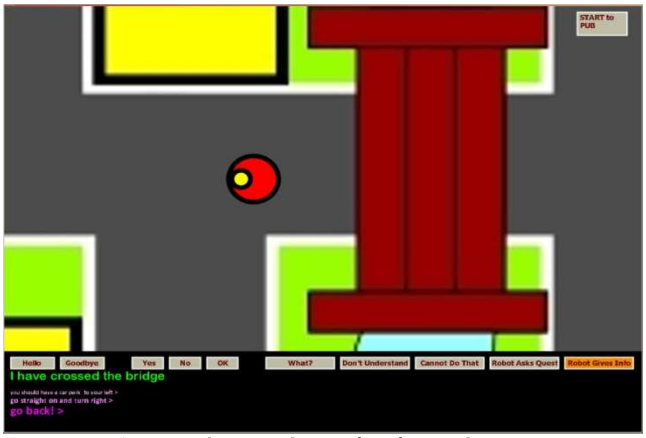


Figure 3. The Wizard's Interface for Condition 2

[2] http://processing.org

[3] These responses signal problems in the utterance, meaning and action level, respectively. See section 2.2, Table 2.

The user had access to the full map of the town (Figure 4), in which the destination was shown in red and the locations that had been reached in previous tasks in blue. On the upper right corner of the screen, the user could see a small fraction of the map that showed the surrounding area of the current position of the robot. Thus, the user could see the surrounding area changing while the robot was moving, but not the robot itself. The dialogue box was on the lower part of the screen. The user could type in messages and send them to the wizard in a manner similar to a messaging application. The messages from the robot were displayed at the bottom of the textbox (in purple). The user's interface did not show the history of the dialogue in order to simulate the transience of spoken language.

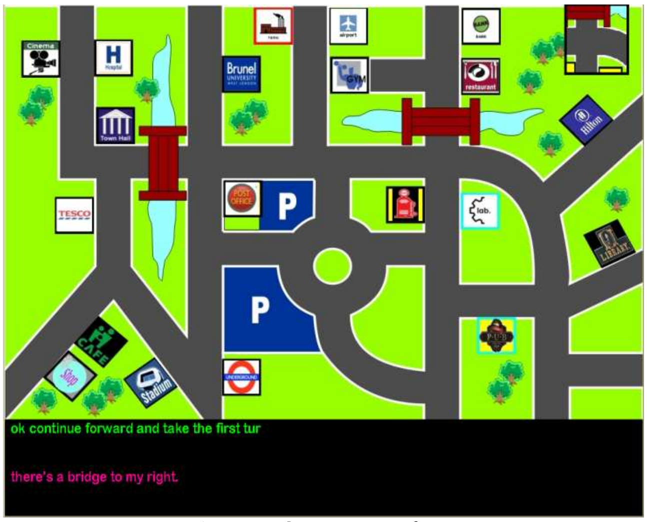


Figure 4. The User's Interface

3. 4 Procedure

Eighteen paid participants were recruited, nine users and nine wizards. Computer expertise was not required. For each experimental condition, three pairs were used. Participant pairs were randomly assigned to one of the three conditions, either as wizards or users. The pairs were seated in separate rooms in front of a desktop PC. The user was sitting alone during the experiment whereas the experimenter was in the same room as the wizard in case they needed technical assistance. The wizards received instructions, a short demonstration and had a trial period to familiarise with the functionality of the system.

The users were briefed about the experiment and the interface. Unlike the wizards, who were aware about the simulation, the users were told that they would interact with a real robot through a computer interface. They were asked to start each task whenever they felt ready by clicking on the links on their computer screen. They were also advised to start each interaction with "Hello", which automatically opened the wizard's application, and end the interaction with "Goodbye" which closed both user and wizard's windows. The user could terminate the interaction at any point by typing "abort task". After each task (completed or aborted), they had to fill in a short questionnaire. The questionnaire consisted of five Likert-scale statements for which the users stated their level of agreement.

Each pair attempted six tasks; specifically, the user guided the robot from a starting point to six designated locations. The destinations were selected in an effort to increase the number or complexity of instructions and the possibility of reference to known routes. Dialogues were allowed to run until the user ended them or up to 10-11 minutes (this cut-off time was decided on the basis of pilot studies).

Both participants received a written description of the experiment which they read before the beginning of the experiment and could consult throughout it. They were not given any instructions on what to say. However, they were explicitly instructed not to use directions such as "north", "south", "up" and "down", but use relative directions from the robot's point of view and common landmarks such as the bridges, the car parks and known locations. Additionally, the users were told that the robot was very fluent in understanding natural language and producing appropriate responses. They were also informed about the route-learning ability of the robot. Last, for the conditions in which the wizards were able to type their own messages, they were asked not to try to "sound like a robot" but talk naturally.

In designing and executing the experiment, a set of parameters as described in [24] were considered: first, the interface which was built for the experiment fully simulated the existing system and allowed for a mixed initiative interaction. Thus, the insights gained from the experiments are relevant to this domain and application while remaining generalisable. Second, the task was open and the focus was not its fast completion, that is, the user could plan or later modify the route in any way. This allowed for dialogue variation. Finally, pilot studies were conducted to ensure that the simulation environment operated by the naive wizard was usable.

3. 5 Experimental Hypotheses

The WOz simulations will be completed in two phases. Phase 1 is described in this paper. Phase 1 aims to generate a refined hypothesis and guide the experiments of the second phase.

The existing system, the robot, (see section 2.1, [3] and [4]), as with the majority of SDS, deals with all sorts of miscommunication in a formulaic way. On the opposite end, human interlocutors have a powerful repertoire of error management skills. Thus, we maintain that a system with more sophisticated error handling capabilities could improve performance. Moreover, previous research has shown that explicitly signalling non-understanding (e.g., "I don't understand" and "Please repeat") is not the default strategy used by people, but implicit strategies (such as clarification requests and providing task-related information) are typically opted for.

The experimental hypotheses are the following:

1. Condition 3, in which the wizard handles miscommunication in an "uninformed", prescribed way, will result in lower success rates and user satisfaction than the "informed" wizard condition (Condition 2).
2. In experimental Conditions 1 and 2, wizards will not resort to explicit signalling of non-understanding.

We also aim to observe several interesting phenomena that are specific to HRI and are not found in HHI and interaction with SDS. These insights, still sparse in literature, will help us develop a system that meets the needs and desires of its user.

4 RESULTS

This section presents an elementary analysis of the data obtained. A total of 54 dialogues were collected and analysed in terms of task success, task completion time, miscommunication, wrong executions and user perceptions.

4.1 Task Success

In Condition 3, one user aborted a task and another pair of participants exceeded the 10-minute time limit and the task was interrupted. In Condition 2, all tasks were completed within ten minutes (see third column of Table 3). This could suggest that in the restricted wizard condition, the participants were unable to recover from errors as effectively as participants in the other conditions. Nevertheless, given the size of the sample, further experiments are needed to validate this claim.

Table 3. Summary of Results from All Three Conditions.

Condit-ion	Ave-rage Time per Task (min)	Task Complet-ion Rate	Miscom-munication Turns/ Total Turns	Total Number of Wrong Executions	% Resolved Wrong Executions	Total Number of No Executions
1	4.58	94.44%	11.29%	10	72.22%	0
2	4.25	100.00%	11.13%	32	77.40%	0
3	5.52	88.89%	12.89%	10	68.75%	56

4.2 Task Completion Time

As shown in the second column of Table 3, Condition 2 resulted in faster interactions compared to the other conditions. The difference between conditions 2 and 3 is greater than the difference between conditions 1 and 2. This might indicate that simplistic signalling of non-understanding leads to longer interactions. Aborted and interrupted tasks were not considered in the analysis.

4.3 Miscommunication and Wrong Executions

Based on the definitions given in Section 2.3, dialogue turns that expressed non-understanding or contained user corrections were labelled as miscommunication. The data in column 4 of Table 3 reveal a high occurrence of such problems across all conditions. In Condition 3, the number of miscommunication turns is marginally larger.

As explained in Section 3.2, the path that the robot followed could be reproduced and each instruction and the corresponding robot action were juxtaposed. If the execution did not match the instruction, it was tagged as wrong execution. As shown in the fifth column of Table 3, the number of wrong executions was significantly higher in Condition 2. However, recovery rates from wrong executions were similar across the first two conditions, but slightly lower in Condition 3. This could imply that misunderstandings that led to wrong executions were more easily resolved in Condition 2 than in Condition 3. It should be also noted that although wrong executions in Conditions 1 and 3 were equally frequent, in Condition 3, 56 "no executions" were tagged (compared to none in Conditions 1 and 2). "No execution" tags were placed in turns in which, although the instructions had been received, there was no reaction (movement) whatsoever by the wizard. It could be speculated that the number of wrong executions could have been higher in Condition 3. But more significantly, this observation supports the idea that the wizards perceived the inadequacy of their expressive means to proactively or reactively deal with problematic understanding. On the other hand, wizards in Condition 1 and 2 felt more confident that they could prevent or resolve wrong executions by signalling non-understanding, requesting clarifications etc. and, thus, were more willing to act based on their assumptions.

4.4 User Perceived Task Success and Overall Satisfaction

After each interaction, the users completed a six-point Likert-scale questionnaire in which they rated their agreement with five statements. These statements covered ease of use, accuracy and helpfulness of the system, perceived task completion ("I think I did well in completing the task") and overall satisfaction ("I am generally satisfied with this interaction"). The responses were mapped to integer values ranging from 1 to 6 (with 6 representing the optimal score). The average scores for perceived task success and user satisfaction were calculated and plotted against each condition (see Figure 6). It seems that, in Condition 2, despite the similar frequency of miscommunication and higher wrong execution rate, the users experienced the dialogue as smoother and more successful.

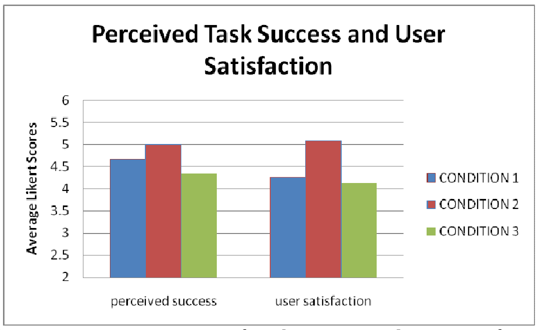


Figure 5. User's Perception of Task Success and User Satisfaction

4.5 Further Observations

Initial turn-based analysis of the dialogues produced several interesting observations. First, in Condition 1, in which wizards were free to formulate their own responses, there was not a single occurrence of explicit problem-signalling responses such as "I do not understand" or "What do you mean?". Even less "talkative" wizards preferred to request further instructions than to state that there was a problem. Explicit strategies should have been the easiest "way out", especially in Condition 2 in which the wizards could simply press a button. However, the "least effort principle" is always balanced and compromised against the need to ensure understanding [5]. Thus, wizards felt that in certain circumstances of interaction, explicit strategies were inadequate. Second, an interesting phenomenon was observed which was not anticipated; wizards in Condition 2 used a balanced combination of explicit and implicit strategies. A sample of a dialogue under Condition 2 is shown in Table 4. In this example, the wizard explicitly signals non-understanding, using the button "I don't understand" (line 2), but then she types a clarification request termed task-level reformulation[4] (line 4).

It can be argued that the physical presence of the buttons in the wizards' interface biased them to use the predefined responses. However, it is equally plausible to assume that these strategies matched exactly what they intended to say. Within the context of the results discussed in Sections 4.1-4.4, this has important implications for the application of the findings of this study to a real-world system. In particular, unconstrained natural communication with a robot, as represented by Condition 1, is infeasible to implement with current technology. However, a system with a smart policy for selecting from a predefined set of explicit and implicit responses, which yields similar benefits in performance and user experience, is a solution within our reach.

Table 4. An Excerpt of a Dialogue in Condition 2.

1.	user said- turn right after the bridge then
2.	robot said- Sorry, I don't understand.
3.	user said- turn right now
4.	robot said - at a car park?
5.	user said- before the car park

Table 5. An Excerpt of a Dialogue in Condition 1.

1.	user said- please turn right and turn right again
2.	robot said- Crossroads, now where?
3.	user said- turn right
4.	user said- keep going right
5.	robot said- I cannot go down that road any more
6.	user said- go forward
7.	robot said- There is a fork in the road
8.	user said- ok turn left here then take the third right
9.	robot said- Before the bridge?
10.	user said- sorry this is the wrong place
11.	robot said- Wrong place? Should Robot go back?
12.	user said- go forward and at the crossroads keep going forward and the tube is at the end of the road
13.	robot said- Out of bounds

Table 5 contains part of a dialogue in Condition 1. In any given situation, the wizard did not explicitly signal the problem but either provided information (lines 2, 5, 7, 13) or used clarification requests (lines 9, 11). In this example, we have the opportunity to observe the balance in the interaction and roles of the interlocutors as they try to cope with and compensate for insufficient sources of information. Management of miscommunication, thus, is a joint activity and not the responsibility of one speaker. The issue of repair rarely becomes the focus of the interaction; the goal of the interaction, that is, obtaining adequate information to complete the task, was not sidetracked but also became a means of repair itself.

Another interesting phenomenon was encountered in Condition 3; as the dialogue progressed, the users tended to pay less attention to the wizard's messages. Instead, they focused on the movement of the robot as they could see it from the little window on their interfaces, and continued issuing commands in a "trial and error" process. This could suggest that the users felt that the robot was no longer helpful or reliable in terms of interaction. This observation is also supported by the post-task questionnaires.

Table 6. An Excerpt of a Dialogue in Condition 3.

1.	user said- go to the lab
2.	user said- do u know where the lab is?
3.	user said- go straight
4.	user said- go left
5.	user said- stop
6.	user said- turn right
7.	robot said- Can you please repeat that?
8.	user said- go straight
9.	user said- go right
10.	user said- move
11.	user said- go straight
12.	user said- go right
13.	robot said- Sorry, I don't understand.
14.	user said- right

Table 6 shows an excerpt of a dialogue under Condition 3. In line 2, the user asks a question but does not wait for a reply. Similarly, the problem-signalling messages of the robot (lines 7 and 13) seem to go unnoticed. The dialogue looks disconnected and the user gives commands and waits to see the action on the screen. Evidently, in real deployment environments such conduct has different consequences.

In conclusion, the numerical results presented in Sections 4.1-4.4 seem to provide support for the first experimental hypothesis. The data discussed in 4.5 only partially confirmed the second hypothesis but revealed a path worth pursuing. Quantitative analysis of the turn by turn unfolding of the dialogue is currently being carried out, using a dialog act tagging scheme. Results are expected soon to illuminate patterns of behaviour regarding wizard strategies and subsequent user responses to address what they perceived to be problematic.

5 DISCUSSION AND FUTURE WORK

The preliminary results reported in this paper are consistent with previous research (see Section 2.3.3) and point towards fascinating research directions. As part of our current and future work, we are running more experiments in order to ascertain that these initial findings are statistically sound and not opportunistic. Moreover, we are performing a fine-grained analysis of the dialogues which involves turn-based annotation of the data with dialogue act tags. The object of the analysis is the actions employed by wizards and users to maintain and restore understanding. These findings will feed the second cycle of simulations and are anticipated to offer additional insights in miscommunication management in HRI. Further, we will look at the relation between wizard strategies, user responses, error recovery, overall dialogue performance, task success and user perceptions.

In [17] and [18], the wizard had no direct access to the user's speech, but could listen to or read the output of a simulated or real speech recogniser. In our simulations we allowed the wizard to have full access to the messages sent by the user without a mid-component. The experimental setup was designed to provide a complex environment that would give rise to many instances of miscommunication (see Table 3) and

[4] In [10], Gabsdil defines task-level reformulations as clarification requests that reformulate the previous speaker's utterance in terms of its effects and the task rather than its surface form.

difficulties in coordination through high referential ambiguity and deictic discrepancies. At this stage in the research, it was decided that ASR would unnecessarily add to the task complexity and negatively affect the consistency and validity of the results; first, the role of the wizard which was assumed by a non-expert participant was already demanding as they had to make decisions and respond fast and accurately with limited and ambiguous sources of information while operating the robot. Dealing with the often incomprehensible output of the speech recogniser would have rendered their task impossible. Secondly, ASR performance is a major source of problems with detrimental effects on the overall system performance. This could confound the effect of the other sources of problems and the error handling strategies, which form the aim of this study, and obscure the observation of other interesting phenomena. The proposed framework will be ultimately implemented and tested using a fully operating system. However, time and resources permitting, we aim to design another experimental condition in which the same set of error handling strategies, as defined by the second round of experiments, is used by the wizard whose abilities are further constrained by ASR.

A valid argument against the experimental setup could be that the results are an artifact of or only relevant to text-based interaction. However, in [11] it is argued that text-based synchronous interaction is now a commonplace means of communication between people (e.g., instant messaging and chat rooms). Moreover, spoken dialogue contains many other sources of information that play a role in participants' understanding. These are extra-linguistic features such as variations in voice amplitude, pitch and speed which function as cues for the listener. When speech is transcribed for analysis these are very hard to represent and are often ignored. On the other hand, text-based interaction constrains users to convey meaning using only linguistic means, which minimises the analyst's manipulation of the data and provides clearer indications of how understanding is achieved. Nevertheless, experiments using real-time speech and audio are planned after these exploratory studies are terminated.

The Map Task offers a rich domain for task-oriented interaction [23]. In the Map Task design, there is one A-role speaker who holds all information necessary for the completion of the task. In the setup of our study, we attempted to raise the status of the wizard; namely, the wizard needed to successfully communicate the surroundings and the exact position of the robot, otherwise the route instructions from the user that were mainly formed with deictic expressions were meaningless. Similar to normal HHI and human cooperative behaviour, in Conditions 1 and 2, participants collaborated and helpfully shared relevant knowledge. Breazeal et al. [25] state that the goal of the field of human-robot interaction is broader than interaction; rather it should be pursued as human-robot collaboration. Thus, the insights and findings that are beginning to emerge from these experiments could contribute in closing the gap between HHI and HRI, so that robots are not tools but partners that play a positive, practical and lasting role in human life.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for their valuable comments and insights.

REFERENCES

[1] S. Thrun, *Toward a Framework for Human-Robot Interaction, Human-Computer Interaction*, 19: 9-24 (2004).
[2] G. Bugmann, Effective Spoken Interfaces to Service Robots: Open Problems, In: *Procs. AISB 2005*, Hatfield, UK, pp. 18–22 (2005).
[3] S. Lauria, G. Bugmann, T. Kyriacou, J. Bos and E. Klein, Training Personal Robots Using Natural Language Instruction. *IEEE Intelligent Systems*, pp. 38–45 (2001).
[4] G.Bugmann, Final Report to the EPSRC http://www. tech.plym.ac.uk/soc/staff/guidbugm/ibl/IGR-IBL.pdf (2003).
[5] H.H. Clark and E. Schaefer, E. Contributing to Discourse. *Cognitive Science*, 13:259–294 (1989).
[6] H.H. Clark, Using Language. Cambridge University Press, Cambridge (1996).
[7] J. Allwood, An Activity Based Approach to Pragmatics. In: *Abduction, Belief and Context in Dialogue, Studies in Computational Pragmatics*, Amsterdam. John Benjamins (1995).
[8] G. Mills and P. G. T. Healey, Clarifying Spatial Descriptions: Local and Global Effects on Semantic Co-ordination. In: *Procs. 10thWorkshop on the Semantics and Pragmatics of Dialogue* (2006).
[9] G. Hirst, S. McRoy, P. Heeman, P. Edmonds, and D. Horton, Repairing Conversational Misunderstandings and Non-Understandings. *Speech Communication*, 15:213–230 (1994).
[10] M. Gabsdil, Clarification in Spoken Dialogue Systems. In: *Procs. 2003 AAAI Spring Symposium Workshop on Natural Language Generation in Spoken and Written Dialogue*, Stanford, USA. (2003).
[11] M. Purver, P. G. T. Healey, J. King and G. Mills, Answering Clarification Questions. In: *Procs. SIGDIAL*, Sapporo, Japan (2003).
[12] D. Schlangen, D. Causes and Strategies for Requesting Clarification in Dialogue. In: *Procs. SIGDIAL*, Boston, MA (2004).
[13] G. Gorrell, I. Lewin, and M. Rayner. Adding intelligent help to mixed-initiative spoken dialogue systems. In: *Procs. 7th ICSLP*, Denver, CO (2002).
[14] B. A. Hockey, G. Aist, J. Dowding, and J. Hieronymus. Targeted help and dialogue about plans. In: *Procs. 40th Annual Meeting of the ACL*, Philadelphia, PA (2002).
[15] D. Bohus and A.I. Rudnicky, Sorry, I Didn't Catch That! - An Investigation of Non-understanding Errors and Recovery Strategies, In: *Procs. SIGDIAL.* Lisbon, Portugal, (2005).
[16] T. Zollo, A Study of Human Repair Initiation Strategies in the Presence of Speech Recognition Errors, In: *Working Notes of the AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems*, (1999).
[17] G. Skantze. Exploring Human Error Recovery Strategies: Implications for Spoken Dialogue Systems. *Speech Communication*, 45(3):207-359 (2005).
[18] J. D. Williams and S. Young. Characterizing Task-Oriented Dialog Using a simulated ASR Channel. *ICSLP*. Jeju, South Korea (2004).
[19] H.H. Clark and C. Marshall, Definite Reference and Mutual Knowledge. In: *Elements of Discourse Processing*. B.L. Webber, A.K. Joshi, and I.A. Sag (Eds.) Cambridge University Press, New York (1995).
[20] N. M. Fraser and G.N. Gilbert, Simulating Speech Systems. *Computer Speech and Language*, 5: 81–99 (1991).
[21] S. A. Green, S. M. Richardson, R. J. Stiles, Billinghurst, and J. G. Chase, Multimodal Metric Study for Human-Robot Collaboration. In: *Procs. ACHI 2008.* Washington, DC (2008).
[22] E. Levin and R. Passonneau, A WOz Variant with Contrastive Conditions, In: *Procs. Dialog-on-Dialog Workshop, ICSLP*. Pittsburg, PA (2006).
[23] G. Brown, Speakers, Listeners and Communication. Cambridge: Cambridge University Press, (1995).
[24] N. Dahlback, A. Jonsson and L. Ahrenberg, Wizard of Oz Studies-Why and How, In: *Readings in Intelligent User Interfaces*, M. Maybury and W.Wahlster (Eds.), Morgan Kaufmann (1998).
[25] C. Breazeal, A. Brooks, D. Chilongo, J. Gray,G. Hoffman, C. Kidd, et al. Working Collaboratively with Humanoid Robots. In: *Procs. International Conference on Humanoid Robots*, Los Angeles (2004).

Video observation of humanoid robot movements elicits motor interference

Aleksandra Kupferberg[1], Stefan Glasauer[1], Markus Huber[1], Markus Rickert[2], Alois Knoll[2], Thomas Brandt[3]

Abstract. Anthropomorphic "humanoid" robots are suggested to be more competent in social communicative interactions than industrial robots, because humans interact more intuitively with them. It is, therefore, critical to evaluate the acceptance of an agent as possible partner for joint interaction. One possible method is to utilize the phenomenon of motor interference (MI). It claims, that observation of an incongruent movement of another individual leads to a higher variance in one's own movement trajectory. Although this effect has been demonstrated while observing a human agent, the researchers were unable to show increased variance in the subjects' movements if they observed an arm of an industrial robot moving with piecewise-constant velocity. In contrast, in other recent studies, MI was demonstrated when subjects watched a humanoid robot performing biological movements based on prerecording of a human experimenter, even if it was not the case when the same robot moved with constant (artificial) velocity. The purpose of the present study was two-fold: 1) we aimed to replicate these results using video-presentations of the agents, and 2) we asked whether quasi-biological movement trajectories are sufficient to elicit MI. We presented subjects, who were instructed to perform horizontal and vertical arm movements, with videos of a human agent or of a humanoid robot, who performed congruent or incongruent arm movements. Robotic movements were produced with a quasi-biological minimum-jerk velocity profile. We found MI both for the human agent and the robot, suggesting that an artificial human-like movement velocity profile (minimum-jerk) is sufficient to facilitate perception of humanoid robots as interaction partners, and that the measurement of MI using a face-to-face video setup can serve as a tool for objectively evaluating humanoid robots.

1 INTRODUCTION

Humanoid robot technology is developing at an incredible rate [1, 2]. In the near future, humanoid service robots equipped with mechanisms for communication and interaction will become part of daily lives of ordinary people. Based on the human's instructions and control, they will assist humans as "partner robots" in completing a variety of tasks that are physically demanding, unsafe, unpleasant, or boring. They will also assist elderly, individuals with physical impairments and cognitive disabilities in care, therapy and training.

During interaction with a humanoid robot, people tend to anthropomorphise it [1] and to apply their experience with human partners in order to explain, understand or predict its behaviour. Since the interaction with humanoids should be natural, enjoyable and efficient, it is important to analyze how we perceive them and how they affect us.

Although the quality of interaction between humans and humanoid robots has been investigated by some studies [2], mostly only questionnaire-based subjective judgments were used for this purpose [3, 4, 5]. A possibly objective tool, which is based on the phenomenon of motor interference, has been developed only recently [6].

Motor interference - influence of the observed movement on own action

By the means of fMRI (functional magnetic resonance imaging), it has been shown, that observation of an action leads to activation of corresponding motor areas in the premotor cortex [7]. It therefore seems that perception of an action leads to simulative production of that action on the part of the observer, facilitating its execution. The neural basis for the "action-perception coupling" hypothesis has come with the discovery of the mirror neuron system in the premotor cortex of macaques, which is activated both when the monkey performs a specific action and when it passively observes the experimenter perform that same action [8,9]. It is presumed that observing the movement of the partner leads to the activation of the premotor areas that correspond to the production of that movement, irrespective of whether the observed movement is compatible with an intended movement. Thus, when the participant observes the partner producing an incongruent movement, the motor program or representation associated with the observed movement is assimilated or interferes with the outgoing motor output for the intended movement. In line with this hypothesis goes the observation that during action observation there is a significant increase in the motor-evoked potentials from the hand muscles that would be used if making such a movement [10]. Thus, observing a certain action injects bias to the motor controller by activation of modules subserving the observed movement *(motor resonance)* and deactivation of modules controlling incongruent movements [11, 12]. This deactivation leads to *Motor Interference* (MI), defined by an increase of variance in one's own movement while watching an incompatible movement. Indeed, it has been demonstrated that

[1]Center for Sensorimotor Research, Clinical Neurosciences, Ludwig-Maximilians-Univ. Munich, Germany. Email: {**akupferberg, mhuber, sglasauer**}**@nefo.med.uni-muenchen.de**
[2]Robotics and Embedded Systems Lab, Dept. of Computer Science, Technical Univ.München, Germany. Email: **{knoll, rickert}@in.tum.de**
[3]Chair for Clinical Neurosciences, Ludwig-Maximilians-Univ. München, Germany. Email: **tbrandt@med.uni-muenchen.de**

while the observation of another person performing an action facilitates the execution of a similar (congruent) action (e.g. the contagion of yawning [13]), it interferes with the execution of a different (incongruent) action [14]. For example, perceiving a horizontal arm movement facilitates the concomitant execution of the same action and curbs the execution of a vertical arm movement [6].

Influence of agency and naturalness of motion on motor interference

It has been shown that although monkey mirror neurons discharge, when it observes an action performed by another individual, these neurons are not activated in the case of mechanical action [15]. However, the question of whether the human mirror neuron system responds to the movements of non-human actors such as robots, is currently controversial. One study [16] suggested that responses in premotor cortex (thought to be the most important part of the mirror neurons network [9]) to observed motion is specific to human action and does not respond to robotic movements. Other data suggest equivalent neural responses to both observed human and robotic action [17].

In the original study investigating robotic movement [6], the MI experimental paradigm was adapted to investigate the extent of similarity of the implicit perception of an industrial robot and a human agent. In contrast to observation of a human agent, MI could not be demonstrated for the observation of movement of an industrial robot. However, some recent studies [18, 19] have found MI when subjects watched a humanoid robot performing movements based on implemented prerecording of a human experimenter. Interestingly, this effect could not be shown when the same robot moved with a constant-velocity profile, suggesting that velocity profiles of biological movements might be essential for MI.

Several studies tried to determine what information in the observed movement triggered the MI in the observer's action. These studies investigated continuous movement synchronization of a human observer with a moving dot stimulus [20, 21, 22], ball motion [23], or a point-light figure [24]. In [22] and [23], the variance of the participant's movements in the incongruent condition was found to be enhanced only with biological motion; in [21], the increase in variance was seen both in biological and in non-biological dot motion; in [24], the biological dot motion led to MI only in subjects with autism in contrast to healthy ones, and in [24], the MI could not be demonstrated neither in human nor in point light figure condition. Although the results of these studies were not uniform, in general they suggest that the brain processes biological and non-biological movements in a different way.

In the experiment reported here, we tried to replicate the results of [18] and [19], by replacing live presentations with video presentations of the human experimenter and the humanoid robot JAST (s. Fig. 1). JAST has an "animal" head and is capable of producing movements with human-like minimum-jerk velocity profiles [2]. The subjects were instructed to produce congruent or incongruent movements while watching the videos projected on a screen.

2 METHODS

Four female and six male PhD students from the local Department of Neurology have been tested in the present experiment. The videos of both JAST and the human agent were rear-projected on a white screen (120cm*160cm) in pseudo-randomized order. The screen was positioned about 1.5 m in front of the participant.

The subjects were instructed to perform 50-cm amplitude horizontal (H) or vertical (V) rhythmic arm movements with their right arms while fixating on the hand of a human agent or JAST (s. Fig 1). The agent performed either spatially congruent (C, same direction) or incongruent (I, perpendicular) movements (frequency: 0.5 Hz). This resulted in a 2*2*2 experiment design with eight experimental conditions and three factors (1) plane of movement (H/V), (2) congruency (C/I), and (3) observed agent (agency; H/R).

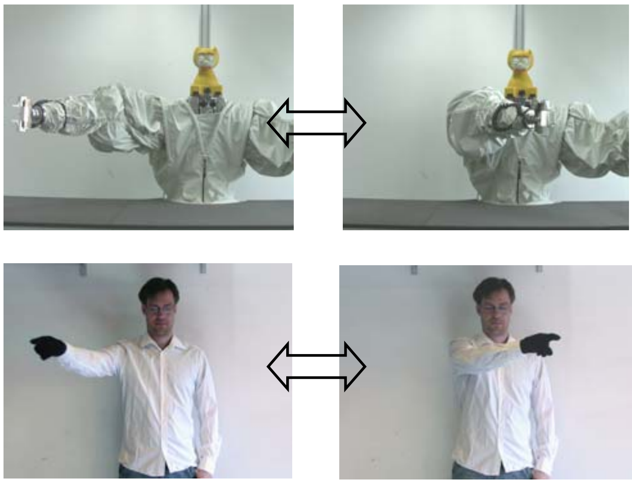


Fig. 1 Screenshots from the videos presented to the subjects. The participants had to make horizontal or vertical movements while fixating on the hand of JAST or a human agent, who performed congruent or incongruent movements.

The robot JAST produced horizontal and vertical movements directed by the shoulder joint. The motion velocity was based on a minimum-jerk profile [25], which, in contrast to the constant velocity profile, makes the movements look smoother and more natural by preventing abrupt changes in movement velocity [2].

One trial (duration: ca. 30s) was performed for each of the eight conditions. At the start of each new condition, the participants were informed (by an instruction appearing on the screen) of the plane in which to move their arm and instructed to keep in phase with the experimenter's and robot's movements. The kinematics of the endpoint of their right index finger was recorded at 240 Hz using the magnet-field based motion tracking system Polhemus Liberty.

After data acquisition, fingertip positions were filtered with a 20-Hz second order Butterworth filter and the data from each trial was split into single movement segments (from right to left and from top to the bottom and vice versa) by finding data points at which the x- and z-values reached their maxima and minima. The standard deviation of fingertip position within the plane orthogonal to the plane of movement was used to quantify the interference. The mean of the deviations of all single movements within one trial was calculated for each subject and then across all the participants.

3 RESULTS

The analysis showed that the mean values (across 10 subjects) of the deviations in the plane orthogonal to the movement were generally bigger for the observation of incongruent than congruent movements for both human and humanoid robot observation (s. Fig 2 for an example of movement trajectories projected onto the x-z-plane). The analysis of variance performed on the factors congruency, plane of movement, and agency, revealed only a main effect of congruency [$F(1,9)=21.8$; $p=0.001$] (s. Fig 3). There were no other significant main effects or interactions between any of the factors.

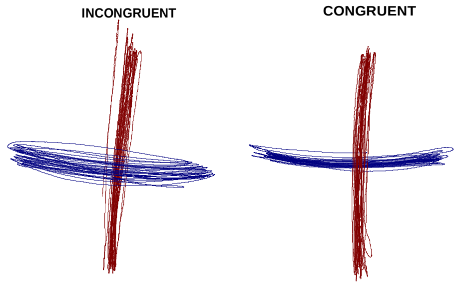


Fig. 2: Individual movements made by a single subject in the XZ- plane. During his movement, the subject observed another human, who performed incongruent or congruent movements.

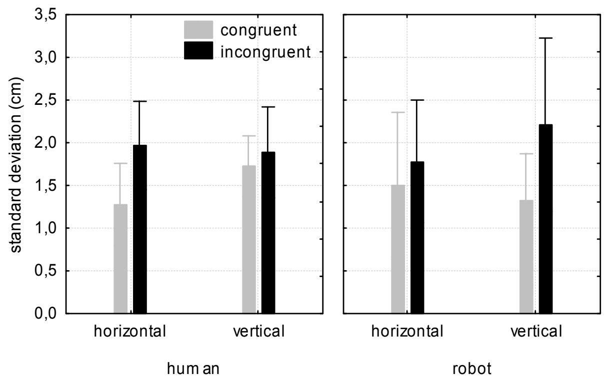


Fig. 3: Standard deviations in the plane orthogonal to the movement plane averaged across ten subjects during observation of a humanoid robot or a human experimenter performing congruent or incongruent movements. Error bars represent 95% confidence intervals.

4 DISCUSSION

While performing everyday activities, individual's limb and body movements are constantly influenced by the observed movements of other individuals [26]. A well-known example of triggering imitation by observing an action is the contagion of yawning [13] by activation of motor systems, which control the observed action. Likewise, the observation of a certain action, different from the own action, can interfere with it.

Movement synchrony enhances intention understanding

Since mirroring the actions of others might help to understand what another person is doing [27], the main function attributed to motor resonance is action perception. Motor resonance might even underlie more sophisticated mental abilities, such as understanding the intentions of others [28]. Thus, simulating another person's actions allows humans to make predictions about the mental states of others based on the mental states and behaviours that they experience themselves while mimicking others [29]. On the neuronal level, it was reported that lesions of premotor cortices involved in the control of action impairs the perception of biological motion presented using point-light displays [30].

The higher degree of movement synchronisation is generally regarded to be a sign of higher degree of mutual rapport, involvement and togetherness [31]. It has been shown, that behavioral synchrony during a dyadic interaction triggers increased attention to the interaction partner leading to enhanced memories about his appearance and his utterances [32]. In psychotherapeutic counseling, congruent movement of limbs of the therapist and the client were significant contributors to attributions of rapport [33]. Additionally, it has been demonstrated, that while asking for route directions for a certain destination, most subject synchronized their arm gestures with the person or the humanoid robot, providing them with these instructions [34]. Another study, investigating body movements in human-robot interaction, has found a positive correlation between the arm movement synchrony of the robot and the human and subjective evaluations of the interaction [35].

Motor interference as objective tool for evaluation of human-robot interaction

As stated in the introduction, motor interference is the direct consequence of motor resonance [13]. Since motor resonance is linked to the sense of togetherness and is observed in a successful human interaction, MI can be used as an objective tool for evaluation of human-robot interaction. Specifically, it can be used to study, what aspects of robot form and motion make it sufficiently human-like and which aspects should be left robotic to display the robot's non-human capabilities. Since the interference effect can be obtained in different planes of movement, the paradigm of MI might be adapted to investigate how naturally other complex robot motions are perceived by humans. However, it might still be helpful to additionally correlate objective findings from the MI experiments with traditional subjective evaluation based on questionnaires.

Present results in light of other studies

The variability of the subjects' arm movements in the orthogonal plane of movement was significantly increased while observing incongruent vs. congruent movements of a human agent (Fig. 2). This was also the case for the humanoid robot. Additionally, there was no interaction effect between the factors "agency" and "congruency". These results support the notion that, during observation of non-goal directed action, the specific neural networks subserving that particular movement are already tuned for action [11], thus interfering with a different action.

The current findings are similar to [18] and [19], who also reported MI for the observation of incongruent action of both human and humanoid robot and strengthen the conclusion, that biological motion velocity might be essential for MI. However, our results extend these previous studies by showing that MI is present in observation of video presentation as well. Furthermore, MI did not depend on accurate biological movement profiles such as used in previous studies: the minimum-jerk movements used in our study, which only approximate biological motion, were sufficient to elicit strong MI.

Together with previous findings [18], our results indicate that the phenomenon of MI is not only limited to observation of human action. A humanoid robot with a limited human-likeness in its appearance may trigger the same type of implicit perceptual processes as a human agent, given that it moves with a quasi-biological velocity. In contrast to that, in the original study [6], MI could not be shown when subjects were observing an industrial robot performing the actions. However, this discrepancy can be explained by the fact that the robot used in the earlier study did not have any humanoid facial features and moved with an artificial constant velocity.

The influence of anthropomorphism on the motor interference

It is traditionally assumed, that building robots with humanoid appearance is the obvious strategy for integrating them successfully into human environments and increasing their acceptance for the majority of non-technical users. However, the question is, whether we need a fully anthropomorphic synthetic human or if a certain degree of form realism is sufficient for social acceptance. A popular theory about the perception of robots [36] states that as a robot increases in humanness, it becomes more susceptible to failures in its functionality and design ("The Uncanny Valley"). This results from the fact that the more human-like the robot appears, the higher are the expectations of people interacting with it. This hypothesis predicts, for example, that a prosthetic limb covered with skin-colored rubber, which imperfectly, albeit extremely closely, reproduces the texture and the motion of real limbs would be more repulsive than a less realistic limb with a mechanical appearance. Therefore, to meet the users' expectations, there must be an appropriate match between physical familiarity with a human and cognitive abilities of the robot.

6 CONCLUSIONS & FUTURE WORK

The present experiment replicates the results of recent studies, claiming that observing incongruent arm movements made by a humanoid robot with a biological velocity, may have a significant interference effect on simultaneously executed human movements. MI also remains stable if the live presentations of the robots are substituted by videos, which are projected on a screen in life-size.

The currently used robot head "iCat" had a zoomorphic appearance with movable eyebrows, eyelids, eyes and lips [37]. This animal-like form might have resulted in the higher acceptance by humans by decreasing the probability of getting into the "uncanny valley, since our expectations of animals' capabilities are lower than of human. The presence of detailed face feature might also have had a positive effect on the emergence of the MI, since it has been shown, that the four features that increase the perception of humanness the most are the eyes, nose the eyelids and the mouth [38].

Initially, MI has been demonstrated in the robot DB, facial features of which are merely suggested, but which, on the other hand, has more degrees of freedom in his joints than JAST and thus a higher capability for biological motion [18, 19]. Therefore, the importance of using a humanoid form in interactive robots is still an assumption that has yet to be proven. Also, the exact aspect of biological motion, which is the trigger for interference, and which is absent in robotic movements (e.g. non-constant velocity, curved trajectory, increased movement variability), remains unknown [39].

Therefore, in the next step, we would like to use the MI paradigm (combined with the subjective evaluations) in order to separate the relative contributions of form and motion to the effect of MI. The question whether humanoid form is essential to elicit a motor response similar to human movement observation can, for example, be investigated by comparing JAST with an industrial robot such as JAHIR [40], which can also be programmed to produce minimum-jerk velocity movements.

Although the MI paradigm is an easy and cheap method for the evaluation of humanoid robots, its applications may be limited to mobile humanoid robots provided with torso and at least two upper limbs. Therefore, it might be helpful to expand the results obtained my MI using additional objective and subjective (questionnaires) methods. Possible objective tools for evaluation of human-robot interaction might include measuring a) physical proximity between the interaction partners, b) number of human approaches towards the robot, c) effectiveness of the jointly completed work, d) success in the solution of mutual tasks, amount of shared attention, e) quality of emotional response to the robot or f) recording heart rate and skin conductivity as measures of arousal.

Together with previous studies, our results will provide a test bed for analyzing human-robot interaction and thus principles for developing guidelines for the future design of assistive robots. These interactive robots will facilitate social competence and support appropriate and pleasant human-robot interaction.

REFERENCES

[1] B. Duffy. Anthropomorphism and the Social Robot. *Robotics and Autonomous Systems* 42:177-190 (2003).

[2] M. Huber, M. Rickert, A. Knoll, T. Brandt, and S Glasauer. Human-Robot Interaction in Handing-Over Tasks. In: *Proc. 17th IEEE Int Symp on Robot and Human Interactive Communication*, pp. 107-112, (2008).

[3] M. P. Blow, K. Dautenhahn, A. Appleby, C. L. Nehaniv and D. Lee,. "Perception of Robot Smiles and Dimensions for Human-Robot Interaction Design." In: *15th IEEE Int Symposium on Robot and Human Interactive Communication (ROMAN 06)* pp. 469-474 (2006).

[4] D. S. Syrdal, M. L. Walters, K. L. Koay, S. N. Woods, K. Dautenhahn. Looking Good? Appearance Preferences and Robot Personality Inferences at Zero Acquaintance". In: *Technical Report of the AAAI - Spring Symposium 2007, Multidisciplinary Collaboration for Socially Assistive Robotics*, pp. 86-92 (2007).

[5] J. Goetz, S. Kiesler, and A. Powers, Matching robot appearance and behavior to tasks to improve human-robot cooperation. In: "*ROMAN 2003. The 12th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2003)"*, pp. 55- 60 (2003).

[6] M. Kilner, Y. Paulignan and S. J. Blakemore. An interference effect of observed biological movement on action. *Current Biology* 13:522–525 (2003).

[7] G. Buccino, F. Binkofski, G. R.Fink, L. Fadiga, L. Fogassi, et al.. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience.* 13:400–404 (2001).

[8] V. Gallese, L. Fadiga, L. Fogassi and G. Rizzolatti. Action recognition in the premotor cortex. *Brain*, 119:593-609 (1996).

[9] G. Rizzolatti and L. Craighero. The mirror-neuron system. *Annual Review of Neuroscience* 27:169–192 (2004).

[10] L. Fadiga, L. Fogassi, G. Pavesi and G. Rizzolatti. Motor facilitation during action observation: a magnetic stimulation study. *J. Neurophysiology* 73:2608-2611 (1995).

[11] M. Jeannerod. The representing brain – neural correlates of motor intention and imagery. *Behavioural Brain Sciences* 17:187– 202 (1994).

[12] W. Prinz. Perception and action planning. *European Journal of Cognitive Psychology* 9:129–154 (1997) .

[13] T. Chaminade. Applying Motor resonance to humanoid robots. In: *Proceedings of IRO'S 2008,* Nice, France, September 26, (2008).

[14] M. Brass, H. Bekkering and W. Prinz. Movement observation affects movement execution in a simple response task. *Acta Psychologica* 106:3–22 (2001).

[15] V. Gallese, L. Fadiga, L. Fogassi and G. Rizzolatti,. Action recognition in the premotor cortex. *Brain* 119:593–609 (1996).

[16] Y. F. Tai, C. Scherfler, D. J. Brooks, N. Sawamoto and U. Castiello The human premotor cortex is 'mirror' only for biological actions. *Current Biology* 14:117-120 (2004).

[17] V. Gazzola, G. Rizzolatti, B. Wicker and C. Keysers. The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *NeuroImage* 31:1674-1684 (2007).

[18] T. Chaminade, D. Franklin, E. Oztop, and G. Cheng. Motor interference between Humans and Humanoid Robots: Effect of Biological and Artificial Motion. In *IEEE 4th International Conference on Development and Learning, Osaka (Japan)*, pp 96-101 (2005).

[19] E. Oztop, D. Franklin, T. Chaminade and G. Cheng. Human-humanoid interaction: is a humanoid robot perceived as a human? *International Journal of Humanoid Robotics* 2:537-559 (2005).

[20] E. Gowen, J. Stanley and R. C. Miall. Movement interference in autism-spectrum disorder. *Neuropsychologia* 46:1060-1068 (2008).

[21] J. Stanley, E. Gowen and R. C. Miall. Effects of agency on movement interference during observation of a moving dot stimulus. *Journal of Experimental Psychology: Human Perception and Performance* 33:915-926 (2007).

[22] C. A. Bouquet, V. Gaurier, T. Shipley, L. Toussaint and Y. Blandin. Influence of the perception of biological or non-biological motion on movement execution. *Journal of Sports Sciences* 25:519–530 (2007).

[23] J. Kilner, A. F. Hamilton and S. J. Blakemore. Interference effect of observed human movement on action is due to velocity profile of biological motion. *Social Neuroscience* 2:158-166 (2007).

[24] S. Jackson, N. Brady, F. Cummins and K. Monaghan. Interaction effects in simultaneous motor control and movement perception tasks. *Artificial Intelligence Review* 26:141-154 (2006)

[25] T. Flash and N. Hogan. The coordination of arm movements: An experimentally confirmed mathematical-model. *Journal of Neuroscience* 5:1688–1703 (1985).

[26] N. Sebanz, H. Bekkering and G. Knoblich. Joint action: bodies and minds moving together. *Trends in Cognitive Science* 10:70-76 (2006).

[27] G. Rizzolatti, L.Fogassi and V. Gallese Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature. Reviews Neuroscience.* 2:661–670 (2001).

[28] M. Iacoboni, I. Molnar-Szakacs, V. Gallese, G. Buccino, J. C. Mazziotta and G. Rizzolatti. Grasping the intentions of others with one's own mirror neuron system. *Library of Science* 3:529-535 (2005).

[29] C. Breazeal, D. Buchsbaum J. Gray, D. Gatenby, and B. Blumberg. Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots. *Artificial Life* 11:31–62 (2005).

[30] A. P. Saygin. Superior Temporal and Premotor Brain Areas Necessary for Biological Motion Perception. *Brain* 130:2452-2461 (2007).

[31] T. L. Chartrand and J. A. Bargh. The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology* 76:893-910 (1999).

[32] C. N. Macrae, O. K. Duffy, L. K. Miles and J. Lawrence. A case of hand waving: Action synchrony and person perception. *Cognition* 109:152-156 (2008).

[33] D. L. Trout and H. M. Rosenfeld. The effect of postural leans and body congruence on the judgment of psychotherapeutic rapport. *Journal of Nonverbal Behavior* 4:176-190 (1980).

[34] T. Ono, M. Imai and H. Ishiguro. A model of embodied communications with gestures between humans and robots. In *Procedings. 23rd Annual. Meeting Cognitive Science Society*, pp. 732–737 (2001).

[35] T. Kanda, H. Ishiguro, M. Imai and T. Ono. Development and evaluation of interactive humanoid robots. In *Proceedings of the IEEE (Special issue on Human interactive robot for psychological enrichment)* 92:1839–1850 (2004).

[36] M. Mori. On the uncanny Valley. In: *Proceedings of the Humanoids-2005 workshop: Views of the Uncanny Valle,* Tsukuba, Japan, September (2005).

[37] F. Hegel, M. Lohse, A. Swadzba, K. Rohlfing, S. Wachsmuth and B. Wrede. Classes of Applications for Social Robots: A User Study. In *Proc. Int. Symposium on Robot and Human Interactive Communication (RO-MAN)*, Jeju Island, Korea, (2007).

[38] C. DiSalvo, F. Gemperle, J. Forlizzi and S. Kiesler (2002). All robots are not equal: The design and perception of humanoid robot heads. In: *Proceedings of the Conference on Designing Interactive Systems* (2002).

[39] S. J. Blakemore and C. Frith. The role of motor contagion in the prediction of action. *Neuropsychologia* 43:260–267 (2005).

[40] G. Reinhart, W. Vogl, W. Rösel, F.Wallhoff and C. Lenz. JAHIR - Joint Action for Humans and Industrial Robots. *Fachforum "Intelligente Sensorik - Robotik und Automation"*, Bayern Innovativ - Gesellschaft für Innovation und Wissenstransfer mbH, Augsburg (2007).

.

The role of expectations in HRI

Manja Lohse[1]

Abstract. Whenever humans interact they form expectations which their behaviour is grounded in. This paper claims that this is also true for human-robot interaction (HRI). It introduces assumptions from expectation theory and illustrates how these can be tested in an HRI context. The paper aims to show that knowing the users' expectations can help to design robots. Since user expectations depend on the robot behaviour, they can be shaped to improve the interaction.

1 INTRODUCTION

The term "expectation" is used in many contexts in every-day life: employees have expectations toward their employers, the society has expectations toward politicians, and someone can expect a baby. Formally, expectations refer to norms, as well as to what happens in the case of uncertainty. Norms are general design guidelines which strongly depend on cultural factors, e.g., distances a robot has to keep from the human (e.g. [1],[2]). Within a culture these norms are rather stable. In contrast, expectations in case of uncertainty strongly depend on the person and on the situation. Based on their own experiences, all people form their own expectations. These are driven by conscious and unconscious memories.

Concerning norms (prior expectations based on culture that users bring into the interaction situation) some studies have been conducted in the context of HRI. These do not use the term expectations but rather refer to attitudes. One study has been reported by Khan already in 1998 [3] when even fewer people than today had interacted with a robot. The findings of the survey with 134 Swedish subjects suggested that certain tasks were preferred for intelligent service robots: robots should help with chores like polishing windows, cleaning ceilings and walls, cleaning, laundry, and moving heavy things. However, they were not wanted for tasks like baby sitting, watching a pet, or reading aloud. 69% of the subjects found the idea of a robot as assistant positive, 44% found it probable, 23% frightening, and 76% found the concept useful [3].

In 2006, Arras and Cerqui [4] have supported the idea that the attitude towards robots depends on the tasks they are built to accomplish. In a survey with more than 2000 participants at the Swiss expo, 71% of the interviewees exhibited a neutral attitude towards robots; meaning that they neither rejected them nor were completely positive. This was due to the fact that the result was based on the mean of attitudes toward robots used for different tasks. As mentioned above, some tasks were very welcome while others were not.

These findings have recently been supported by Ray and colleagues [5] in their publication titled "What do people expect from robots?". In a questionnaire study on people's perception of robots with 240 participants, the authors found a positive attitude towards robots in general. But they also reported that people prefer robots mainly as helpers in daily life and not as caregivers in whatever context. They noted that people expressed their fear about robots replacing humans and about robots becoming too autonomous. The authors' findings revealed that robots should not be humanoids or look like living beings. On the other hand, while appearance should not be human-like, people prefer natural channels, e.g., speech, for communication with robots.

Attitudes have also been in the focus of research with the NARS (negative attitudes towards robots) scale, which is based on the concept of computer anxiety (e.g. [6], [7]). However, researchers using this scale barely identified a link between attitudes and behaviour. In this sense, knowing one's attitude does not allow for a reliable prediction of actions because the behaviour is also strongly influenced by the situation [8]. In general, a positive attitude towards robots cannot lead straight to the conclusion that people in fact use the systems. Therefore, I state that a priori expectations and attitudes are not the main determinants of individual interactions. Instead, it is the situation that mainly affects the expectations during the interaction. Just as situations, in my approach expectations are seen as dynamic. This assumption is based on a constructivist model of mind (see Section 2B). Similar to memories, expectations need not only be recalled but they are actively constructed in a situation. Thus, the situation strongly influences what comes to a person's mind (accessibility of an expectation), which schemas are used for interpretation explicitly or implicitly (explicitness of expectations), how a person interprets what is occurring and which attribution processes take place (how the person logically justifies the occurrences). Therefore, I assume that changes of expectations can be observed during the interaction. This approach attempts a more reliable explanation of user behaviour in HRI than attitudes and a priori, normative expectations because it generalizes less and takes into account the particularities of the situation and online changes of expectations during the interaction.

The first part of the paper introduces expectation theory, revealing different kinds of expectations (Section 2). Section 3 provides some thoughts on the importance of the situation in which an interaction is taking place. In Section 4 assumptions based on the theory are summed up. Thereafter, I present a case study which shows how the assumptions might be tested based on data from user studies (Section 6). The data used for the analysis is described in Section 5. Based on the analysis, I discuss the relevance of expectations in HRI (Section 7).

[1] Applied Informatics, Research Institute for Cognition and Robotics, Bielefeld University, Bielefeld Germany.

Email: mlohse@techfak.uni-bielefeld.de.

2 EXPECTATION THEORY

Before taking a closer look at expectation theory, I want to clarify why in this paper both the terms "expectation" as well as "expectancy" are used. I agree with [9] that in the literature it seems that both termini appear interchangeably. I prefer "expectation" but if literature is cited, I use the same term as the authors. The discussion of a possible distinction between the terms is not in the scope of the current paper.

In a recent paper, Roese and Sherman [10], give an overview of expectancy theory. They define expectancies as follows:

> "Expectancies are beliefs about a future state of affairs, subjective estimates of the likelihood of future events ranging from merely possible to virtually certain." (p.91)
>
> "The expectancy is where past and future meet to drive present behavior" (p.92)

In the first version of the article from 1996, Olson, Roese, and Zanna [11] call the kind of expectationcies defined here *probabilistic expectancies* since they describe what could happen with a certain probability. The definition presented above indicates that past experiences shape expectancies. On the other hand, expectancies support anticipating the future and, therefore, shape the behaviour of a person.

One classic experiment in social psychology which shows the importance of expectations was conducted by Kelley (1950) (see [12], [13]). In the experiment, students were told that a guest instructor was either warm or cold. After the lecture students judged the instructor. Their judgments clearly showed assimilation effects: the group of students that were told that the instructor was warm hearted, rated him as being more considerate, informal, sociable, popular, good-natured, humorous, and humane. These students were also more open for a discussion with the instructor. In this example expectations primed in the students beforehand influenced information processing. Based on this experiment one could assume that participants enter situation with expectations which do not change over time. In contrast to the experiment in most situations, and also in our experiments, however, people are not primed for specific expectations beforehand. How expectations form if they are not primed is described in the following section.

A Sources of Expectations

Olson and colleagues [11] name three major sources of expectations: direct personal experience, communication from other people (indirect experience), and beliefs that are inferred from other beliefs. Personal experience determines whether expectations are *semantic* or *episodic*. According to Roese and Sherman [10], semantic expectancies are

> "preexisting knowledge structures that are extracted from ongoing experience, stored in memory, and retrieved when needed. By contrast episodic memories must be formulated on the spot before they can be applied" (p.95).

While semantic expectancies are an efficient way of using past experiences, episodic memories provide depth by not including abstract knowledge but information from the situation. Semantic and episodic expectancies interact and can be applied in parallel. Usually, people at the outset have specific expectations, which become more general over time, and are finally balanced at mid-levels. Based on this concept, one can assume that expectations change over time. Especially in situations in which users interact with a robot for the first time changes are probable.

In 1977, Heckhausen [14] has introduced a model that also points out the influence of the situation on expectancies. The model encompasses a four-stage sequence of events:

- the initial situation as perceived
- the person's action
- the outcome of the action or the situation
- the consequences of the outcome

Based on these stages, Heckhausen names different kinds of expectancie: situation-outcome expectancies, action-outcome expectancies (these two are typically addressed in motivation theory, e.g., by Tolman), action-by-situation-outcome expectancies, and outcome- consequence expectancies.

Maddux [15] applies two different terms (behaviour outcome expectancy and stimulus outcome expectancy) which seem to be similar to action-outcome expectancies and situation-outcome expectancies, respectively. Behaviour outcome expectancies are beliefs that a specific behaviour results in a specific outcome or set of outcomes. Stimulus outcome expectancies "are concerned with beliefs that certain events provide a signal or cue for the possible occurrence of other events" ([15], p.23).

The next section provides some more details about the function of expectations which is closely connected to their processing.

B Function and Processing of Expectations

> " […] people think about the future and use such thoughts in ongoing judgment, reasoning, decision making, and behavior." ([10], p.91)

This statement shows that expectancies influence behaviour. They provide an effective way to guide behaviour by providing a shortcut in mental processing. Judgments about the future are not stable but rather construed by the human. In the construction process expectations guide information gathering as top-down processes influencing event processing; and they provide structure and meaning for the interpretation of the gained information. On the other hand, they can also lead to inaccurate inferences if they are incorrect, if they bias information collection, or if they overrule consideration of information altogether [16].

In accordance with these findings, Roese and Sherman [10] state that expectations influence information seeking. Sometimes people try to validate their hypothesis of the world (at least for subjective expectancies) and search for more information. Validation processes occur especially in cases of expectation disconfirmation. Disconfirmation is identified with a mechanism that Roese and Sherman term regulatory feedback loops. According to the authors, the current state of a situation is compared to an ideal or expected state. If discrepancies are discovered, the person will search for behaviour in order to correct or reduce the discrepancies.

> "Behavior control therefore requires continuous online processing in the form of continuous comparison, or pattern matching, between the current state and the expected state. Very likely in parallel to this comparative process is the conceptually similar online comparison between the current state and recent past states." ([10], p.93)

In most cases expectations are confirmed since behaviour is constantly based on simple expectations, e.g., if we push a light-switch we expect that the light comes on (or goes off if it was on beforehand) because we have experienced this many times before. The confirmed expectations are usually processed heuristically. Then again, heuristically processed information is more likely not to bring to awareness disconfirmation. Disconfirmation of expectations causes more careful processing. The expectation is perceived as being less certain. Disconfirmed expectations become more explicit and accessible and are therefore more likely to be tested in the future. Next to these basic assumptions of expectation disconfirmation, Olson et al. [11] list some more cognitive consequences of disconfirmation:

- both consistent and inconsistent information receive more processing than irrelevant information
- people prefer to interpret information in line with their expectancies
- unexpected events trigger attributional processing but attributions also influence expectancies (reciprocal relationship)
- disconfirmation triggers more counterfactual processing, but expectancies also shape the semantic content of counterfactual thought (reciprocal relationship)
- expectancies provide a cognitive structure to more easily encode and retrieve consistent information from memory, on the other hand by means of expectancies unexpected events become salient which increases memorability

Olson et al. [11] conclude that behaviour is consistent with the content of expectancies, e.g., people chose tasks in which they expect to succeed and put more effort into them. Expectations can also lead to self-fulfilling prophecies (e.g., [11]). Next to behavioural also affective and physiological consequences (e.g. placebo effect) can be found. Affect usually is negative when expectancies are disconfirmed an exception being, e.g., a surprise party where the expectations are disconfirmed in a positive way.

3 IMPACT OF THE SITUATION

In Section 2A, the model by Heckhausen which links expectations and situation has been introduced. Moreover, it has been discussed how expectations are construed in the situation and how the situation determines what comes to the mind of a person. In this section, I aim at further clarifying how the term situation is understood in this paper.

In contrast to personality traits, situation is a dynamic concept [8]. Many experiments have shown how situational factors superpose personality in human-human interaction; e.g., in Libermann, Samuels and Ross (2004) (in [17]) students played the same game which was introduced to two groups with two different names (Community Game and Wall Street Game). Students in the second condition (Wall Street Game) displayed significantly more competitiveness independent of how their competitiveness was rated beforehand. They obviously constructed a certain idea of the situation. The awareness of the situation was affected by the name of the game.

Situation awareness (SA) as a concept has been formulated by Endsley and colleagues [18]:

> "The formal definition of SA is 'the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future'" (p.13; from Endsley, 1988)

Endsley differentiates three levels of situation awareness: perception (Level 1), comprehension (Level 2), and projection of future states (Level 3). The definition shows that SA is closely connected to expectations in that the person analyzes the surroundings to predict future occurrences and to adjust the behaviour. According to Endsley, expectations, next to goals, mental models, and schemas form the internal representation of state. They influence perception, attention, and other cognitive processes. These, in turn, influence expectations. Therefore, expectations are bound to the situation and to how it is perceived.

In the example by Liberman and colleagues introduced above, the researchers supposed that all subjects conceptualized the situation in the same way based on the name of the game that they played. This is not necessarily true for HRI situations. Fischer [19] has analyzed experimental data for cues how the subjects conceptualized situations. She found that beliefs about the communication partner show in linguistic behaviour and situation influences the language used.

> „the speakers have been found to constantly define and, if necessary, redefine what they understand the situation to consist in, depending on their current hypotheses about their communication partners and on their own emotional state." ([19], p.7)

This work emphasizes, once again, the importance of the situation as such. However, while Fischer is interested in linguistic phenomena, I focus on strategies that include different modalities and what they tell us not only about the perception of the situation but also about users' expectations. Since processing is an online mechanism, I expect that expectations change during interaction, and that the changes are based on what the robot does and what happens in the situation as such.

4 ASSUMPTIONS BASED ON EXPECTATION THEORY

Summarizing the previous sections, I assume that:

1. expectations are construed in the situation and, therefore, may change during the interaction
2. expectations are influenced by the robot behaviour, and the situation.

3. expectations are influenced by the human characteristics (e.g., appearance, age)
4. first contact situation results in not having concrete former expectations → people form episodic expectations and repeat successful strategies
5. information gathering is based on expectations, what people pay attention to (look for, ask) displays what they expect
6. users try to verify their expectations especially after expectation disconfirmation
7. expectation disconfirmation can lead to changes in affect
8. behaviour is consistent with the content of expectations, e.g., people chose tasks in which they expect to succeed and put more effort into them

These assumptions provide an extensive framework for research on expectations in HRI. Not all of them can be tested in this paper. However, all the assumptions will have to be tested if we want to show the influence of expectations on HRI. In the following I introduce a case study as an example on how to determine expectations based on a data-driven approach.

5 HRI DATA

The relation between HRI and expectations can only be researched based on interaction data which is situated in a certain situation because (1) real users can be expected to have different expectations and (2) as has been discussed before, the situation determines the expectations. The test bed in this paper is data from a study in a home tour situation. The home tour focuses on multi-modal human-robot interaction in which the robot should learn about the domestic environment and its artefacts, the appearance and location of objects, and their spatial-temporal relations. With these abilities the robot can adapt to new settings like a user's home. The home tour focuses on the tutoring of the robot. The exploitation of the acquired knowledge and representations for, e.g., conducting fetch-and-carry tasks is a consequent extension of this scenario. Robots' capabilities for natural interaction in the home tour comprise understanding of spoken utterances, co- verbal deictic reference, verbal output, referential feedback, and attention to persons. Furthermore, the robot must be able to follow a human when she is walking around and showing an apartment.

Robot platform

The robot used for the study is called BIRON (BIelefeld Robot companiON). BIRON is based on a Pioneer PeopleBot platform. As can be seen in Fig. 1, BIRON employs various modalities: it can see faces with the camera on top and gestures with the camera below the screen, it displays a character called Mindi on the screen (see Fig. 2, 3, and 4), has microphones for speaker localization (speech recognition is done using a headset), speakers for speech output, and a laser scanner to explore the environment. Details about architecture, design principles and abilities of BIRON can be found in [20].

Subjects

14 subjects participated in the user study. All participants were German native speakers and interacted with BIRON in German. They were recruited at a public event of the university. Their age ranged from 16 to 71 years (average 45.5 years). 9 participants were male, 5 female.

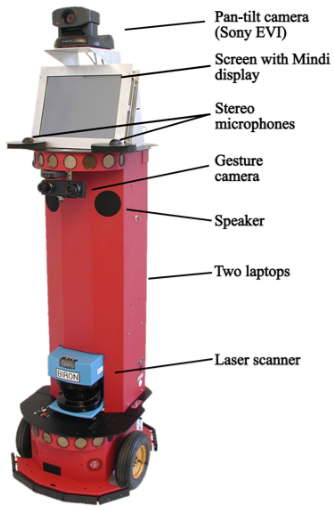


Fig. 1 The robot BIRON

Procedure

On arriving, the participants were welcomed and introduced to the study. They answered a questionnaire on demographic data and their experience interacting with robots. Afterwards, the subjects were trained on using the speech recognition system, i.e., they were instructed about the proper placement of the head set microphone and were asked to speak some phrases for habituation. The recognition results were displayed in verbatim on a laptop computer. Thereafter, participants were guided into the room where the robot was waiting ready for operation. Subjects were assisted during the first contact in order to reduce hesitant behaviours. They were handed a tutorial script for practise. The script contained all commands the users would need later on and the autonomous robot operated just as it did after the training session. During this initial tutorial session the experiment leader also instructed the users on how to pull the robot to alleviate cases when the robot gets stuck. After the tutorial session the training script was handed back and the main task, namely to teach areas and locations in several rooms, was carried out by the subjects. The instruction for this main task was:

- guide the robot through the apartment, i.e. from the living room to the dining room via the hall
- show and label the living room and the dining room
- show the cupboard in the living room and the floor lamp in the dining room

During this part of the interaction, the experiment leader only intervened when prompted by the subjects. The whole interaction was videotaped.

Afterwards, the subjects were interviewed about the interaction in general and the visual display in particular. The interviews have also been digitized. Moreover, the participants answered a second questionnaire which included items on liking of the robot, attributions made towards the robot and usability of the robot. More information about the study is provided in [21].

Data Analysis

The video data of the study was used for the analysis presented here. The corpus was annotated using ELAN. A coding scheme for human speech, gestures, and other movements was developed and employed by three coders. The robot activities were logged automatically. All data (video, manual annotations, and automatic logs) were synchronized in one ELAN file per user in (see [21] for more information).

For the analysis presented here, one situation was chosen from all recordings. The situation is characterized by BIRON having an incomplete percept of the user. The percept consists of a voice and visual features, namely legs and face. If the person perception is unreliable, e.g., because lightening conditions are poor or because the person has not said something in a while, the robot is not able to match the percept to one person.

This state is communicated on the display with a picture of a the screen character Mindi (Fig. 2). Due to the state the robot does not react to the user until the she has done something to improve the perception. In general, solutions to this problem include two strategies: verbal behaviour or movement of the user in front of the robot. The user has the initiative. Since this state is not part of the explicit dialog, it is only communicated using the Mindi display. The situation perception of the users and their expectations determine how they react to the situation. Therefore, the situation is of special interest here.

The situation was chosen because it was an unexpected event in the interaction. In fact, it constitutes a change in the interaction that should also lead the user to adapt her expectations to the new situation. The modified expectation should then result in a change in user behaviour.

In the video data of the 14 subjects, 26 sequences were identified. In the interactions of users 10 and 13 no such sequences occurred. All other recordings contained one to three sequences.

In the most basic form, the sequences are characterized by the following actions. First, the Mindi display changes to "poor person perception" (Fig. 2), the person reacts to the display, the display changes to the "processing" Mindi (Fig. 3), then to the "robot is facing user" Mindi (Fig. 4), and finally the robot says "hello". The average length of these sequences is 13.8 seconds (minimum 4 seconds, maximum 45 seconds; measured from Mindi appearing to Mindi changing to processing).

Apart from the video recordings, the questionnaire that the subjects answered after the interaction was taken into account for the analysis. The questions of interest are related to whether the participants had seen the "poor person perception" Mindi (Fig. 2) and what they thought it meant.

Another aspect that was included in the analysis was that five subjects had received an introduction to the meaning of the picture during the interaction. In the explanation BIRON told them: "If you see this picture, I cannot see or hear you very well. Please change your position and repeat what you have said." Only five people listened to the explanation because the subjects were free to chose whether they wanted to listen to it or not. In Section 7, it will be discussed whether receiving an introduction or not had an impact on the interaction.

Moreover, users 5 and 6 received help from the experimenter when the Mindi in question was displayed on the screen during the training phase. User 5 was told to restart the robot (by saying "restart"). User 6 was asked to say "hello" again in one occasion and to restart the robot in another.

In the training the experimenter at times interrupted the interaction and the subjects often continued with the training script regardless the reaction of the robot. Therefore, only sequences from the main task were included in the analysis.

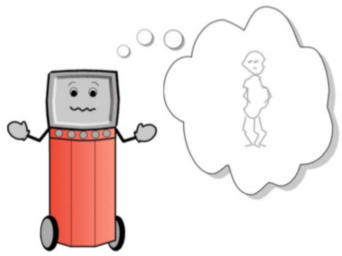


Fig. 2 "Poor person perception" Mindi

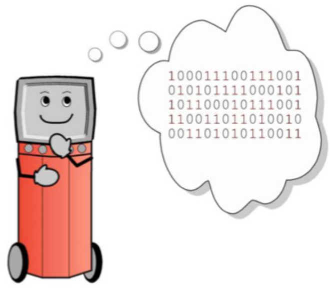


Fig. 3 "Processing" Mindi

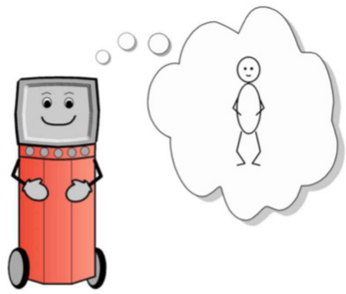


Fig. 4 "Robot is facing user" Mindi

6 RESULTS

I want to anticipate that in fact the new situation changed the expectations of the users. However, there were differences in how they conceptualized the change.

One aspect that all analyzed sequences had in common was that the users searched a position in front of the robot before starting another behaviour to resolve the situation. Users did not only aim to stand in front of the robot body but also tried to be faced by the camera on top of the robot. This could be seen in the videos because all users kept walking around and leaning to the side until the robot faced them. Therefore, the poor perception display triggered the same expectation in all user that the robot could perceive them better when facing them.

Another behaviour that was shared by all participants was that the time users wait before the next action (verbal or movement) was very stable within subjects. However, it strongly differed between subjects. Some only waited two seconds while others waited ten seconds and more for BIRON to give a feedback. The expected feedback could be verbal but also appear on the screen.

The results clearly showed that when the Mindi display changed to “processing”, the users did not take any more actions and waited for more robot feedback.

Another expectation that was shared by all participants was that some verbal utterance was necessary to regain the robot’s attention. However, three different strategies to resolve the situation could be identified:

1) verbal behaviour only
2) verbal behaviour first with movement added
3) movement first with verbal behaviour added

Their distribution is depicted in the following table.

Table 1 Strategies applied by subjects

user	1	2	3	4	5	6	7	8	9	11	12	14
strategy 1	3	3	1	2		1		2	1		2	
strategy 2						1	1	1			1	1
strategy 3				1	3				1	1		

Strategy 1 was characterized by verbal behaviour only. Out of the 26 analyzed sequences, 13 were identified with this group. They mainly occurred in short sequences in which it sufficed that users said something once (either “hello” or something that starts another action like “follow me”); e.g. as depicted in Fig. 5 (time code in the figure caption tells when the sequence occurred in the interaction). The figure illustrates occurrences over time. In the upper row actions of the human are labelled, in the lower row actions of the robot. Red (dark colour) actions are verbal actions while light-grey actions are non-verbal actions (movements of the human, changes on BIRON display).

The average length of these sequences was 7.2 seconds (minimum 4 seconds, maximum 16.5 seconds). In general the sequences were shorter than 8 seconds. The average time people waited after the last action was 6.2 seconds (minimum 3.25 seconds, maximum 7.5 seconds). Measured from the appearance of the Mindi it was 4 seconds (minimum 1.5 seconds, maximum 6 seconds). All these numbers do not include user 12 because she took significantly longer than all other subjects and was obviously distracted by some event that was not connected to the interaction.

In addition to the sequences which were resolved after giving a verbal command once, this group also includes two situations (two different subjects), in which BIRON did not react after the first utterance and the subject only repeated the verbal utterance without any movement.

Strategy 2 contains five cases of five users all of which apart from this strategy used strategy 1 in other sequences. The rest of their sequences were short sequences as described above (saying “hello” or another command once and the interaction continued). In strategy 2, sequences contained saying “hello BIRON” or “BIRON hello”, movement, and at least one more verbal command (Fig. 6). Interestingly, the movements of the users almost only consisted of stepping backwards and forwards in front of the robot (4 instances). Only one person stepped to the left or the right (1 instance). Moreover, in all sequences the exact same wording was repeated. The average length of sequences in this group was 18.25 seconds (minimum 8.75 seconds, maximum 24 seconds).

Strategy 3 includes six cases, three of which by user 5 (all her sequences) and one each of three other users. The strategy was characterized by the user moving after positioning herself in front of the robot and before taking any verbal action.

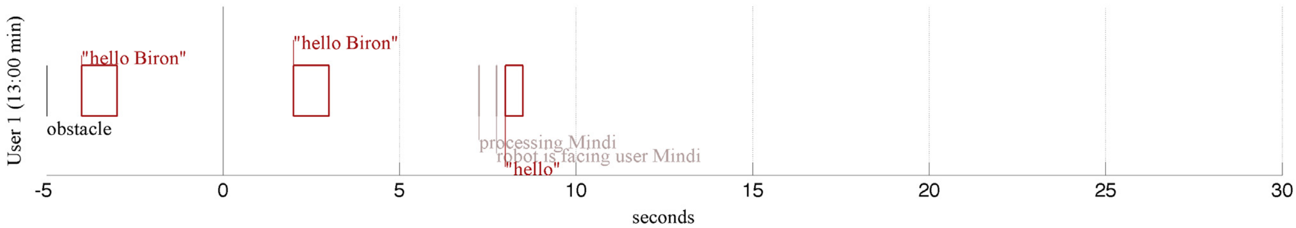


Fig. 5 User 1 (13:00 min)

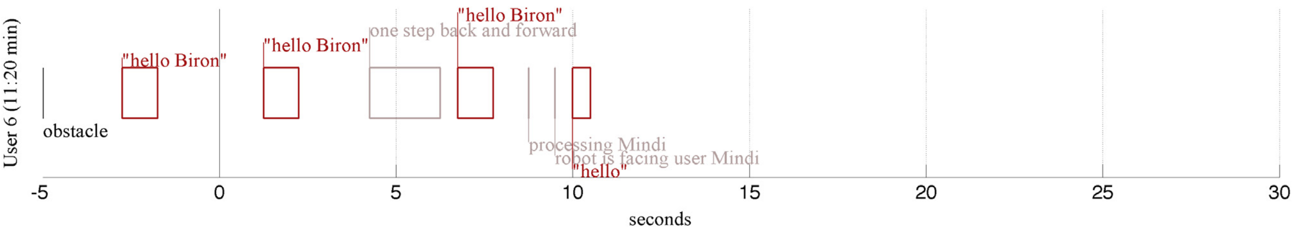


Fig. 6 User 6 (11:20 min)

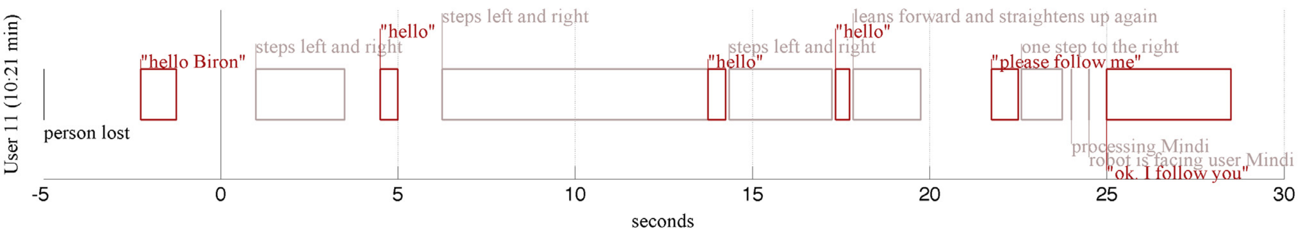


Fig. 7 User 11 (10:21 min)

Altogether, the pauses between movements were shorter than between speech. User 11 actually kept moving almost continuously with only four breaks in a sequence of 24 seconds (Fig. 7).

The average length of sequences with this strategy was 26.5 seconds (minimum 8.25 seconds, maximum 45 seconds) which is much longer than in Strategy 1 and 2.

In contrast to strategy 2, movement to the side is more common here (to the side - 20 times, backward/forward - 2 times, legs apart - 2, times lean forward - 3 times). Only one person in one sequence stepped forward and backward (user 5, 14:06). User 4 (twice in one sequence) and 11 (once) lean forward and position their faces in front of the camera.

Next to the video data we also analyzed the questionnaires to find out why different strategies were employed. Of interest here is how the users interpreted the Mindi in question (Fig. 2). They were asked this as an open question ("What does this picture express?) which they answered in writing. The following descriptions were given (answers were translated by the author):

User 1	BIRON has not understood "hello", is waiting for commands
User 2	BIRON cannot see / hardly hear the user
User 3	BIRON does not see
User 4	BIRON can hardly see the user
User 5	BIRON tries to recognize the user
User 6	BIRON has lost sight of the user
User 7	BIRON does not know / has not understood
User 8	BIRON does not find the person
User 9	the user is unknown/ not recognized any more
User 11	BIRON does not know
User 12	BIRON does not see / does not know what to do with a command
User 14	BIRON does not know what to do in a situation

While nine answers (1, 2, 3, 4, 5, 6, 8, 9, (12)) actually reflected poor perception of the users by the robot, others referred to knowledge and understanding of the robot (7, 11, (12), 14). In most cases, user perception addressed the robot's ability to see (apart from users 1, 2, 7). The answers of users 10 and 13 are not included because their video data did not contain sequences that were analyzed.

The question which arises from the different strategies that can be seen in the videos and the answers from the questionnaires is how users conceptualized the situation and what expectations they formed. It will be discussed in the following section.

7 DISCUSSION

In the following, I want to elaborate on how the findings from video data and questionnaires can help to test assumptions regarding expectations in HRI.

The results showed that, in the situation analyzed here, it was crucial that all users felt that the robot's camera was facing them. This expectation arose in the situation based on the robot's screen output and missing verbal feedback. Based on their experience users new that something went wrong in the interaction. All users reacted to the display with verbal and non-verbal behaviour which signalled that they took over initiative and asked for the robot's attention. Thus, the robot feedback triggered the expectation that the user had to take the turn.

In the sequences analyzed, it was noticeable that all participants accepted the Mindi changing to processing as feedback signal. As soon as the display switched, the users then expected the robot to have the initiative and did not take any more action. Therefore, the display seems to indicate that the robot is doing something and some action will follow. On the other hand, for the poor person perception Mindi the expectation did not seem to be quite clear and users conceptualize the situation differently. This became obvious in two observations: the time that users waited before the next action was quite stable within subjects while the time between users differed; and participants' behaviour in this situation can be grouped in three strategies (verbal behaviour only, verbal behaviour enriched with movement, and movement enriched with speech).

Strategy 1 proved to be the most successful one in that the sequences were shortest and the problem was resolved fastest. Verbal input by the user allowed the robot to identify the interaction partner and to continue with the task. In accordance with expectation theory, users repeated this successful strategy when the same situation reoccurred. Only two people gave the robot verbal feedback also if it failed at first. Their expectation seemed to be that BIRON needed a verbal percept which was probably furthered by many requests of the robot to say "hello" if something went wrong (e.g. after moving the robot away from an obstacle, after having lost the percept of the person completely).

Users that applied strategy 2 also seemed to agree with the importance of greeting the robot to attract its attention even though the questionnaire data showed that their understanding of the Mindi picture differed. Only user 2 did not greet BIRON but started a new action (following). He also was the only one who moved during the utterance, all others only tried one action (speak, move) at a time. His movement was closely connected to the follow command. Therefore, with reference to the situational constraints, the function seems to be to start walking in the right direction rather than to attract attention. What is common to strategy 2 is that verbal utterances were repeated and enriched with movement. The direction of the movement in most cases was forward and backward.

In contrast people applying strategy 3 moved to the left and to the right. This movement was accompanied by a camera movement to both sides. Therefore, it was an obvious feedback that the robot was still doing something. This might have triggered subjects to try out more movements. No camera movement occurred when people walked backward and forward. Therefore, in sum, people applying strategy 2 might have tried more verbal behaviour.

Non-verbal behaviours in strategy 3 also included that user 4 (twice in one sequence) and 11 (once) leaned forward and positioned their faces in front of the camera. They probably assumed that the robot had difficulties to perceive the face but needed this percept in order to continue interaction. This movement was not tried out right away but rather late in the sequence of user 11 and only in the third sequence of user 4.

Strategy 3 was usually only used once. Only user 5 repeated the strategy which was not successful with regards to the time needed to resolve the situation. However, it is noticeable that in the later situations she used a verbal utterance much earlier after the Mindi appeared. Obviously her expectations had changed at least in part. This finding supports the assumption, that the history of the interaction influences expectations.

Concerning the question why people apply a certain strategy, one possible explanation is the training participants received from the robot itself and hints from the experimenter which influence their perception of the situation. As described in Section 5, five subjects had received the introduction to the meaning of the Mindi picture during the interaction. Their behaviour was different to the other subjects in that five of six cases of strategy 3 were used by people who received BIRON's instructions. The verbal introduction by BIRON seems to stress more strongly that the robot has a problem "seeing" than the Mindi picture does. However, it is noticeable that user 11, who received the introduction, interpreted the picture as "BIRON does not know", even though there was another picture with a question mark in the speech bubble which was clearly recognized as BIRON does not know / does not understand.

These finding support the assumption that expectations are influenced by the robot's behaviour. Therefore, robot behaviour should be designed to shape users' expectations and behaviour to enable them to more efficiently solve tasks in the interaction. In the data reported here such an improvement regarding efficiency could not be observed because (1) as the questionnaire entry of user 11 shows users cannot be expected to perceive all information equally well, (2) too few situations were analyzed, and (3) even if more data would have been available it can be doubted whether it would have revealed the improvement because, as the analysis has shown, the feedback in this situation is obviously not yet clear enough to shape users' expectations. However, now having analyzed the data, we know about the implications the robot behaviour has on user experience and can strive for a more helpful feedback. The example, moreover, shows that data acquired in user studies can be used to infer expectations that users have during the interaction. Moreover, it can be assumed that these expectations change based on the situation and on how it is conceptualized by the user.

8 Future Work

This paper is only a start in research regarding expectations in HRI. It introduces assumptions on user expectations which need to be tested to fully explore the influence of expectations on HRI. Based on a case study, it claims that expectations can be inferred from data acquired in interaction situations and that interpretation of data can profit from taking users' expectations and conceptualizations of the situation into account. Knowing the users' expectations can help to design robots because expectations depend on the robot behaviour and can therefore be shaped to improve the interaction.

Another approach that in the future might drive robot design is to not only research user expectations but to also explore robot expectations. Answering the questions "What kind of input does the robot expect?" and "How can it communicate which input is expected?" might also improve the interaction considerably.

Finally, the question on how expectations can be measured more reliably deserves further consideration. The case study presented here revealed some expectations in a very restricted situation. Since these findings should be compared to other interaction situations all of them should be using the same measurements.

REFERENCES

[1] E. Pacchierotti, H.I. Christensen, and P. Jensfelt. Human-Robot Embodied Interaction in Hallway Settings: a Pilot User Study. ROMAN 2005. pp. 164-171 (2005).

[2] M.L. Walters, K. Dautenhahn, K.L. Koay, C. Kaouri, R. Te Boekhorst, C. Nehaniv, I. Werry, D. Lee. Close encounters: Spatial distances between people and a robot of mechanistic appearance. In: Proceedings of the IEEE-RAS International Conference on Humanoid Robots (2005).

[3] Z. Khan, Attitudes towards intelligent service robots. Technical Report Nr. TRITA-NA_P9821, IP-Lab-154 IPLab, Royal Institute of Technology (KTH), Sweden (1998).

[4] K.O. Arras and D. Cerqui. Do we want to share our lives and bodies with robots? A 2000-people survey. Technical Report Nr. 0605-001 Autonomous Systems LabSwiss Federal Institute of Technology, EPFL (2005).

[5] C. Ray, F. Mondada, and R. Siegwart. What do people expect from robots? *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nice, France. pp. 3816-3821 (2008).

[6] T. Nomura, T. Kanda, and T. Suzuki. Experimental Investigation into Influence of Negative Attitudes toward Robots on Human-Robot Interaction. Presented at the *3rd Workshop on Social Intelligence Design (SID2004)*, Twente (2004).

[7] T. Nomura, T. Kanda, T. Suzuki, and K. Kato. Psychology in Human Robot Communication: An Attempt through Investigation of Negative Attitudes and Anxiety toward Robots. *RO-MAN 04* (2004).

[8] K.J. Rohlfing, M. Rehm, and K.U. Goecke. Situatedness: The interplay between context(s) and situation. *Journal of Cognition and Culture* 3(2) pp. 132-157 (2003).

[9] N. T. Feather. Introduction and Overview in: N. T. Feather (ed.) *Expectations and Actions: expectancy-value models in psychology.* Lawrence Earlbaum Associates: Hillsdale. pp. 1-14 (1982).

[10] N.J. Roese and J. W. Sherman. Expectancy. In: A. W. Kruglanski and E.T. Higgins, (ed.) *Social Psychology. Handbook of Basic Principles.* Guilford Press. pp. 91-115 (2007).

[11] J. Olson and N. Roese and M. Zanna. Expectancies. in E.T. Higgins and A. Kruglanski (ed.) *Social Psychology: Handbook of Basic Principles*. Guilford Press: New York. pp. 211-238 (1996).

[12] M. Biernat. Standards and expectancies: contrast and assimilation in judgments of self and others. *Psychology Press* (2005).

[13] E. Aronson, T.D. Wilson, and R.M. Akert. *Social Psychology* Pearson Education; Upper Saddle River, NJ (2003).

[14] H. Heckhausen. Achievement motivation and its constructs: A cognitive model. *Motivation and Emotion* 1 (4). 283-329 (1977).

[15] J. Maddux. Expectancies and the Social-Cognitive Perspective: Basic Principles, Processes, and Variables. In: J. Kirsch, I. (ed.) *How expectancies shape experience*. American Psychological Association. pp.17-39 (1999).

[16] D.O. Sears, L.A. Peplau, J.L. Freedman and S.E. Taylor. *Social Psychology*. Prentice Hall: Englewood Cliffs, NJ (1988).

[17] G. Bierbrauer. *Sozialpsychologie*. Kohlhammer, Stuttgart (2005).

[18] M.R. Endsley, B. Bolté, and D.G. Jones. *Designing for Situation Awareness. An Approach to User-Centered Design*. Taylor & Francis: London and New York (2003).

[19] K. Fischer. What is a situation? *Gothenburg Papers in Computational Linguistics*. 00-05:85-92 (2000).

[20] M. Hanheide and G. Sagerer. Active Memory-based Interaction Strategies for Learning-enabling Behaviors. Proc. International Symposium on Robot and Human Interactive Communication (RO-MAN) (2008).

[21] M. Lohse, M. Hanheide, K. Rohlfing, and G. Sagerer. Systemic Interaction Analysis (SInA) in HRI. *Proceedings of Human-robot interaction conference* (2009).

Iterative design process for robots with personality

Bernt Meerbeek[1], Martin Saerbeck[1], Christoph Bartneck[2]

Abstract. Previous research has shown that autonomous robots tend to induce the perception of a personality through their behavior and appearance. It has therefore been suggested that the personality of a robot can be used as a design guideline. A well-defined and clearly communicated personality can serve as a mental model of the robot and facilitate the interaction. From design perspective, this raises the question what kind of personality to design for a robot and how to express this personality? In this paper, we describe a process to design and evaluate personality and expressions for products. We applied this to design the personality and expressions in the behavior of a domestic robot.

1. INTRODUCTION

1.1 Context

Traditionally, robotic technology has been used in industrial settings, for example in car manufacturing. However, more and more robots appear in the domestic area. In the near future, robots will provide services directly to people, at our workplaces and in our homes [1]. Application areas include household tasks (e.g. vacuum cleaners), security tasks, entertainment purposes (e.g. toys), and educational purposes. While nowadays a technical explanation of how an appliance works is give to the user, this will become increasingly difficult with new autonomous robots. We cannot expect users to learn about sensors, actuators and control architectures. Instead, we need to convey a mental model that helps the user to make sense out of the robot's behaviors and to understand which actions are needed from his side. The appropriate design of the interaction between humans and robots will be a crucial factor for the understanding and acceptance of new robotic products [2]. A promising approach in this field of Human-Robot interaction [3] research is to equip robots with life-like and social characteristics. Fong, Nourbaksh, and Dautenhahn [4] present an overview of what they call socially interactive robots, i.e. robots that exhibit human-like social characteristics. Some examples of these characteristics are the ability to express and perceive emotions, to communicate with natural language, to establish and maintain social relationships, to use natural cues in verbal and non-verbal behavior, and to exhibit distinctive personality and character.

1.2 Animacy and anthropomorphism

It is known from earlier research [5],[6],[7] that robots, will induce the perception of being life-like and having a certain personality, through their appearance and behavior. Heider and Simmel [8] already demonstrated in 1944 that people attribute motivations, intentions, and goals to simple inanimate objects, based solely on the pattern of their movements. Tremoulet and Feldman [9] showed that even the motion of a single featureless dot is enough to convey the impression of animacy. They concluded that animacy is inferred when observable aspects of the display cannot easily be explained as ordinary inanimate motion. Recent field tests, such as the ethnographic study with the robotic vacuum cleaner Roomba conducted by Forlizzi et al. [1], revealed that already the use of an autonomous robot in a social environment (i.e. the home) had an impact on social roles and cleaning habits of the participants, even if the robot was not in particular designed for social interaction.

The cognitive process of attributing life-like features is also known as anthropomorphism (in case one attributes human-like characteristics) or zoomorphism (in case one attributes animal-like characteristics). One of the most debated topics is whether designers should use anthropomorphic features in robots. For example, Ishiguro argues that robots that imitate humans as close as possible serve as an ideal interface for human [11]. Duffy, on the other hand, puts this view in perspective, arguing that anthropomorphic features have to be carefully balanced with the available technology in order to not raise too high expectations that cannot be met [12]. He stresses that the goal of using anthropomorphic features is to make the interface more intuitive and easy to use and not to copy a human. Up to now, the question to what extend to incorporate anthropomorphic artifacts remains unanswered. In line with Duffy, we believe that anthropomorphic or life-like features should be carefully designed and aim at making the interaction with the robot more intuitive, pleasant, and easy. In the next section, we explain how the concept of personality can be helpful in designing appropriate life-like features in a robot.

1.3 Personality

Reeves and Nass [6] have demonstrated with several experiments that users are naturally biased to ascribe certain personality traits to machines, to PCs, and other types of media. For a product designer, it is therefore important to understand how these perceptions of personality influence the interaction and how a coherent personality can be utilized in a product.

Personality is an extensively studied concept in psychology. As McAdams and Pals [13] point out, there is no "comprehensive and integrative framework for understanding the whole person". Carver and Scheier [14] give an impression of the diversity of research on personality. They present an overview of personality theories categorized along seven perspectives, including the biological, psychoanalytic, neo-analytic, learning, cognitive self-regulation, phenomenological, and dispositional perspective. In outlines, these theories agree on the general characteristics of personality, amongst others that personality is tied to the physical body; helps to determine how the person relates to the world; shows up in patterns (recurrent and consistent); and is displayed in many ways (in behavior, thoughts, and feelings).

As our work concentrates on the expression of personality as a pattern of traits, personality research on dispositional traits was considered most relevant. This dispositional perspective is based on the idea that people have relatively stable qualities (or traits)

[1] Philips Research HTC 34 WB51, 5656AE, Eindhoven, NL

Email:{**bernt.meerbeek,martin.saerbeck**}**@philips.com**

[2] Dept. of Industrial Design, Univ. of Tech. Eindhoven,

5612AZ Eindhoven, Email: **c.bartneck@tue.nl**

that are displayed in diverse settings. Dryer [15] stresses three focus points to maintain the coherence of the characters personality: (1) cohesiveness of behavior (2) temporal stability (3) cross-situation generality. A combination of several trait theories that focused on labeling and measuring people's personality based on the terms of everyday language (e.g. helpful, assertive, impulsive, etc.) led to an emerging consensus on the dimensions of personality in the form of the Big-Five theory.

The Big-Five is currently the theory that is supported by most empirical evidence and it is generally accepted [13]. It describes personality in five dimensions: extraversion, agreeableness, conscientiousness, neuroticism, and openness to new experiences. Table 1 provides a comprehensive overview of the five dimensions and some of their facets. These facets indicate the scope of each dimension and the variety of aspects within a dimension. Recent studies have used personality theories such as the Big Five to assess people's perceptions of robot personality (e.g. [16][17]). However, the Big-Five theory of personality can also be used as a framework to describe and design the personality of products, and in particular of robots. Norman [5] describes personality as: 'a form of conceptual model, for it channels behavior, beliefs, and intentions into a cohesive, consistent set of behaviors.'. Although he admits this is an oversimplification of the complex field of human personality, the statement indicates that deliberately equipping a robot with a personality, it helps to provide people with good models and a good understanding of the behavior.

Personality dimension	**Personality facets**
Extraversion	warmth, gregariousness, assertiveness, activity, excitement-seeking, positive emotion
Agreeableness	trust, straightforwardness, altruism, compliance, modesty, tender-mindedness
Conscientiousness	competence, order, dutifulness, achievement, striving, self-discipline, deliberation
Openness	fantasy, aesthetics, feelings, actions, ideas, values
Neuroticism	anxiety, angry hostility, depression, self-consciousness, impulsiveness, vulnerability

Table 1 Five-factor model: dimensions and facets

1.4 Research questions

In sum, appropriate design of the human-robot interaction is an increasingly important research topic as robots move into domestic settings. The important questions that arise when explicitly designing a personality for a robot in a given application are what kind of personality is appropriate for the robot and facilitates the interaction, and how to express the personality in the behavior of the product? Our research investigates how personality can be addressed in the design process.

2. PREVIOUS WORK

Careful design of robotic behavior appears to be a crucial factor for the acceptance and success of a robot application. However, up to now there is no consensus on general design rules for character design, nor a unified design process. Several approaches have been reported on the process to design personalities for expressive autonomous products. In this section we summarize some of the existing approaches relevant to character design. Traditionally, there have been three main perspectives on designing the expressive behavior of a robotic product: (1) technology driven, (2) artistic design (3) user centered. We illustrate each of these approaches next.

2.1 Technology driven

When the first robots were constructed, the behavior was fully determined from a technological, functional point of view. The behavior was implicitly implemented by engineers, who had the technological insight to control the hardware. Hence, the behavior resulted from functional requirements such as navigating via the shortest path to a certain location, as well as hardware constraints such as maximum speed or correction movements for compensating hardware inaccuracies. Several architectures for designing the behavior of interactive robotic characters have been proposed [17][19]. In the subsumption-architecture proposed by Brooks [20], the overall behavior of the robot is explicitly an emergent feature, composed from simpler basic actions.

How the user perceives a certain behavior had only later been taken into account. For example, Kawamura et al. [21] stressed the necessity for ease of use of a service robot, but bases multiple design decisions on technical constraints of a particular robotic platform. Loyall [22] presents a complete architecture to construct autonomous and believable agents that encompassed among others a specialized language to describe the behavior for believable agents.

Neubauer [23] takes a more analytical approach to design artificial personalities. He explores the application of Carl Jungs theory of personality in design of artificial entities such as chat bots or avatars on the web. He classifies personalities according to the classification scheme of Jung and categorizes them according to what personality type is implementable with a computer, given our current understanding of artificial life.

The main characteristic of these approaches is the focus on specific technical implementations. Even though the underlying technology is an essential factor for the feasibility of a robotic application, they tend to narrow the design space by technical limitations, rather than by user insights.

2.2 Artistic design

In contrast to a technical approach, the artistic approach is mainly concerned with the expression of a behavior. The focus is not on the functionality of the robot, but on how people perceive the behavior. The underlying idea of conveying messages through expressive behavior is borrowed from the field of movies and animations. The most cited set of design guidelines are the 12 design principles of Disney Animation by Thomas and Johnson [24] listed in Table 2.

1. Squash and Stretch	7. Arcs
2. Anticipation	8. Secondary Action
3. Staging	9. Timing
4. Straight Ahead Action and Pose to Pose	10. Exaggeration
5. Follow Through and Overlapping action	11. Solid Drawing
6. Slow In and Slow out	12. Appeal

Table 2: 12 Animation principles of Disney animations by Thomas and Johnson

The design principles serve as a tool that focuses on creating believable expressions and behavior in short sequences of a movie. The overall personality of the character is determined by a central movie script.Van Breemen [25] was one of the first to apply animation technology to the development of robots. He illustrated that by simply adhering to some of the animation principles, the behavior of a robot appears to be more life-like.

In general, however, approaching the design of robotic behavior from an artistic point of view requires good artistic skills of the designer. Several guidelines have therefore been developed that support the designer to make and justify choices, but they do not take away the need for creativity and inspiration. Dautenhahn, for example, refers to comic design and identifies two design dimensions: (1) universal design (2) abstract design [26]. On the first dimension, the designer abstracts out universal features of behavior or an expression, so that people can recognize and identify themselves with the character. On the second dimension the designer has artistic freedom to add specific features that can best be described by an artistic style.

2.3 User centered

In the process of investigating design rules for interactive robotic characters, many of the design principles have been borrowed from the field of human-computer interaction [27]. The user centered approach is characterized by a strong focus on the user. The key principle is an iterative design cycle to evaluate and refine the interface. One of the most cited references for design principles in human-computer interaction is Gould et. al [28]. The three proposed principles are: (1) early focus on users and task (2) empirical measurement (3) iterative design. The first principle focuses on understanding the user and task, by having close contact with the user. One suggestion for learning from the users is to use interviews. These initial interviews should be constructed before the first design prototype. The second principle demands to carefully investigate how people interact with the device at hand. The authors warn the designers not only to present a system to the users but also measuring usability data. The assumption for the third principle is that it is almost impossible to get a system interface right the first time, hence promoting an iterative design cycle. Lately, however, these principles were target of critique [29]. A main point in the argumentation was that the success of their design could not be attributed to these principles but was founded on more general design principles.

Many more user-centered design approaches have been reported in literature. For example, Ljungblad et al. [30] surveyed participants that own exotic pets to investigate in what kind of forms and roles of characters people are interested. They used the concept of personas [31] to guide their design process for creating personalities for artificial agents. From the interviews they generalized use cases and scenarios, pointing out that the interviewed persons are not necessarily the intended users of the system.

The notion of designing and validating scenarios rather than focusing on personalities for character design also proved to be useful for designing a personality for the personal robot PaPeRo [32]. The scenarios construct a basic set of interactions with the user, placed in the context of an application. During validation the authors found that due to different colors of the robots, users attributed different personalities and roles. For example, the blue PaPeRo was perceived as the leader of the other PaPeRos, and the yellow one was perceived as if it were the youngest. This feedback was taken into account by changing the behavior to enhance these role perceptions e.g. by changing the utterances of the robot. This way, gradually a personality of the robots could be designed.

Despite the focus on the user, the creative element of the designer still plays an important role in the design process. Friess examined real world practice of a design process and found that during everyday interaction not only usability evidence is used to defend design decisions, but very often also pseudo evidence and simply common sense [33]. Höök [34] proposes a user-centered process and applies it to three case studies. She investigates how affective user interfaces can be designed and how they can be evaluated. She criticizes formal approaches of user-studies, since they do no capture the fine grained facets of personality and affective design. She proposes a two layered design approach. The first level focuses on the usability, by verifying whether basic design intentions such as emotional expressions are understood by the user. On the second level, it is verified whether affective aspects in the design contribute to the experience of the user. The user becomes an integral part of the design process, but instead of formal evaluation of the system, the user should be able to give a broad interpretation of his or her experience. Furthermore, she points out that traditional user studies search for an average user that does not exist. Instead of generalizing, affective design should focus on how the individual interacts with the system.

3. PERSONALITY DESIGN PROCESS

Although several approaches to design personalities for expressive autonomous products have been proposed, we miss a practical process that integrates a user-centered, artistic, and technical approach to designing personalities. In this section, we describe the process that we followed to design personality and expressions for a domestic robot and propose this as a way to design personality and expressions for autonomous products in general. The process consists of five main steps, namely creating a personality profile, getting inspiration for the expressions, sketching a scenario, visualizing it in 3D animation, and evaluating it using a think-out-loud protocol. The focus of this section will be on the general process, rather than on the application specific results.

3.1 Create a personality profile

In the design process we propose, we use the notion of personality as a central design guideline to create consistent and understandable behavior (mental model), to facilitate natural (social) interaction, and to make the product more appealing. Therefore, the first question that needs to be addressed is what kind of personality should be designed for a robot.

We used a user-centered approach to create a personality profile for the domestic robot. As a starting point, we used the most widely accepted personality model in psychology ("Big-Five", see section 1). For each personality dimension, we selected several traits (i.e. personality characteristics) to be used as triggers for potential end-users to talk about the desired personality of a product.

Many questionnaires based on the Big-Five are available (both commercial and non-commercial), which typically consist of a large number of items [35],[36],[37]. Several authors have used single adjectives instead of phrases as personality descriptors [38], [39]. All questionnaires are used to assess one's personality.

However, in our case we want to use the items to get people to talk about particular aspects of the product personality.

We decided to select a subset of traits from all the questionnaires that we reviewed. We selected those items that we expected to be useful triggers for obtaining qualitative feedback from users on the desired personality for the robot, given the applications and tasks of the domestic robot. Furthermore, we made sure that we selected at least two positive and two negative markers for each of the five dimensions. In total, we selected 21 items, since using more items would lead to an unacceptably long session with our participants.

The traits were presented on cards to participants (potential end-users of the domestic robot) and they were asked to explain to the interviewer what the characteristics would mean for the behavior of the robot. Next, they were asked to place the cards with personality characteristics on an A0 sheet to indicate how desired this characteristic was for their preferred robot (see Figure 1). An example: A participant was shown the card with the word ‘polite’ (agreeableness). She explained that this could mean that “when the robot wants to move in the same direction as you do, it will wait and let you go first.”. “Yes, that is a desired behavior. I put it close to the center.” This method resulted in detailed qualitative and quantitative feedback on the personality for a domestic robot. For each trait, the percentage of participants that considered it to be either undesired, neutral, desired, or ideal was calculated. Furthermore, the rationale (why something was desired or undesired) was recorded and analyzed. The subjective rationale provided more insight into what kind of robot behavior people prefer and therefore addressed aspects of the application that have not been anticipated before. Also, users’ gave many examples of robot behavior for each of the presented personality characteristic that on the one hand yield insights on how users interpret robotic behavior and on the other hand can be used to narrow the design space for prototyping of behaviors.

Based on the user feedback, a descriptive personality profile was created. This profile is a narrative description of about 300 – 400 words illustrating the character of the robot. This personality profile can be used in a similar way as personas [31]. While personas are often used to describe users in the target group and communicate it to a development team, the personality profile describes what (‘who’) the product is. This profile provides a frame of reference for later stages in the development of the product behavior.

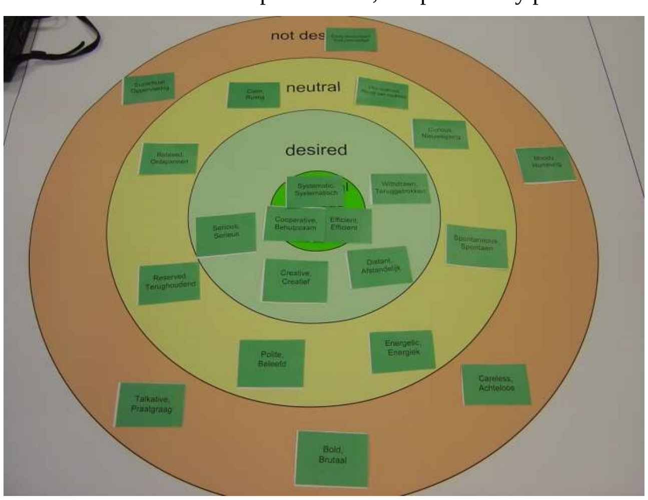


Figure 1 User-created personality profile

3.2 Get inspiration for expressions

In order to get ideas and inspiration for designing life-like and expressive behavior for robots, a theatre workshop was organized. During the workshop, four actors from an improvisational theatre group acted out possible behavior of a domestic robot on the basis of the personality profile. We used the workshop to explicitly address the creative and artistic aspect of the design process for a robotic application. Acting out the behavior of a robot makes sense because it helps to build a basic understanding of the personality. This method has proved successful in acting for movie and theatre for decades [40]. This holds especially for emotional expressions due to the interrelated nature of emotion experience, emotion expression and readiness for action [41]. Experience and expressions reinforces each other. For example, expressing a smile when feeling sad will cheer you up [42]. Translating these results to a design process offers the possibility not only to design emotions in a top-down approach but also bottom-up. The actor workshop was held in a realistic living room setting in the Home area of the ExperienceLab facility at the High Tech Campus in Eindhoven [43]. The session was recorded with 4 ceiling-mounted cameras and one mobile camera. First, the actors studied the personality profile to identify with the character. After that, the actors showed behavior (focusing on its movements and sounds, but no talking) of the robot in particular situations in various ‘exercises’ that are commonly used in improvisational theater. The moderator of the workshop presented a situation to the actors, e.g. ‘You are being switched on’, ‘You are exploring the room’, ‘You encounter an obstacle.’. The actor that had an idea how to act in this situation stepped forward and acted out the robot behavior. The scene ended with a buzzer sound made by the moderator, after which the next actor could show his/her expressions. In another exercise, one actor freely acted out some behaviors, while another actor had to give ‘live commentary’ what he or she was seeing. Some scenes were played individually and some in teams.

Over 200 scenes were recorded. Video cards ([44]) were used to group, compare, and analyze the large amounts of video material (see Figure 2). The clustered video cards with descriptions of the behaviors and example video clips were discussed in the project team. During these discussions, additional ideas for expressions were generated.

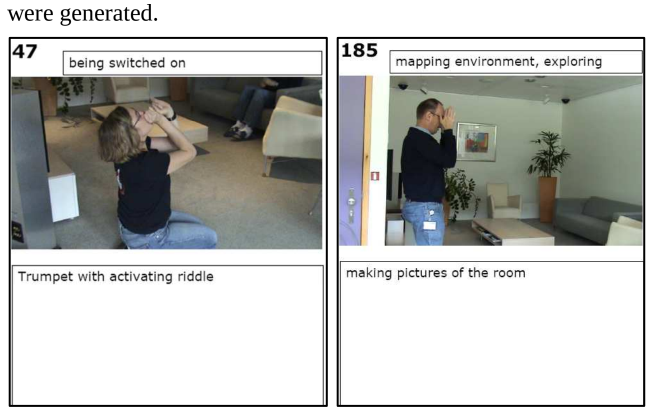


Figure 2 Two examples of Video Cards.

3.3 From actor to robot expressions

The video clips with the expressions of the actors were translated into expressions for the domestic robot. Since human expressions cannot be mapped one-on-one to expressions of the robot, we abstract the human expressions first before we could design concrete expressions for the domestic robot. For example, an actor was looking around and pretending to make pictures of the room to express that he was exploring the environment. This was translated as repetitive turns of the robot to the left and to the right ('looking around'), flashing white lights ('camera flash light'), and a click sound ('picture taken').

The designed expressions were sketched in a written scenario and an animated storyboard. This scenario and storyboard was used to communicate the expressions within the project team. Although presentation of animated behavior on paper is difficult, people inside the project team were able to give initial feedback on the cartoon-like drawings showing the robot behavior. The final storyboard served as input for the visualization of the behaviors in 3D animations.

3.4 Visualize in 3D animation

We used virtual 3D graphical simulations for prototyping and testing scenarios of robotic applications because 3D simulations offer the designer the possibility to present a concrete instance of a particular behavior or scenario to the participants, and gather feedback from users without the hassle of building a fully functional hardware prototype. In product development, designers commonly use sketches and cartoons to visualize certain concepts. While sketches can only show a static representation, a 3D simulation gives an impression how a dynamic behavior will look like. The timing of movements and behaviors is a crucial element for the meaning of an expression [45].

Nowadays, several software packages are being used to simulate robotic behavior, each with their own advantages and disadvantages [46]. In general, robotic simulations can be approached in two ways, either by simulating the physical properties of the hardware and control software, or by scripting the behavior and allowing for artistic freedom without real-world constraints. From our studies we learned that for a first impression, animation technologies are sufficient, because a designer can make a behavior reasonably realistic, while focusing on conveying a message. Animation technologies offer the designer the freedom to implement certain behaviors to give a life-like impression of the robot [25]. However, a close resemblance of the real hardware behavior proofs useful during later stages in the design process, because behaviors can be ported to the real hardware for experiments. Although simulations give a realistic impression of the behavior, research has shown that there are important subtle differences in the perception between virtual and physical embodiment [48].

We used the Open Platform for Personal Robotics (OPPR) framework as described in [49] to develop visual impressions of the robotic behavior in a realistic setting and recorded these in several movie clips. One particular strength of the OPPR framework is that it uses physical simulations for rendering animated behavior. Therefore, the virtual simulation closely resembles the real hardware platform, so that the behaviors can be developed and tested with users and at a later stage reused on a real robot.

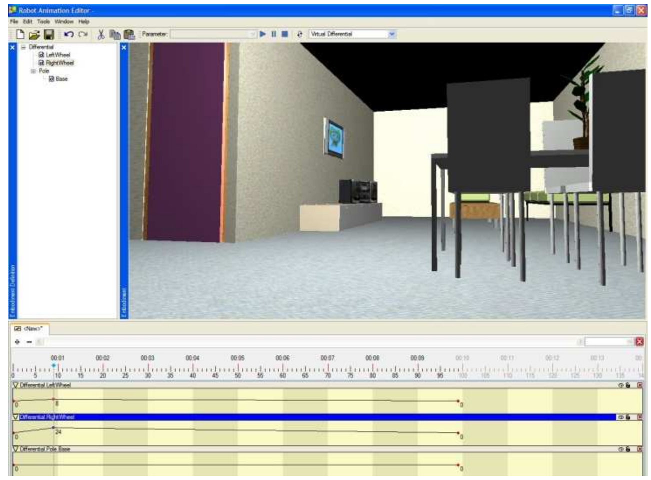


Figure 3 Screenshot of Animation Editor

3.5 Evaluate with think-out-loud

The 3D animations were used in a think-out-loud evaluation. The objective of the evaluation was to find out how people perceive the expressions of the robot and obtain user feedback as input for redesign and/or implementation of these behaviors. Questions we were interested in were: How do people interpret the designed behaviors? What behaviors do people consider life-like and why? Why is some behavior preferred over other behavior?

In total, 12 participants were invited to participate individually in a session of approximately one hour. A video clip of about 8 minutes was shown to the participants with animations of the expressive robot behavior. While watching the video clip, participants were asked to continuously describe what they saw, thought, and felt and why. What do you think the robot is doing or what is it trying to tell you? After the video clip, participants were asked to fill out a questionnaire and were interviewed about their impression of the robot. Audio and video recordings were made of the sessions to allow for verbal protocol analysis. Remarks of the participants were clustered and analyzed per segment of the video clip (typically a segment consisted of one behavior).

The think-out-loud evaluation and verbal protocol analysis resulted in valuable qualitative user feedback on the designed expressions. The results clearly indicated which expressions could be easily interpreted and which expression were appreciated by participants. The results are used as input for the next iteration of designing the expressions.

With the think-out-loud evaluation, we finished the first iteration of our design cycle. We have a clear view on the desired personality for the product and gathered user feedback on the designed behavior. During the next steps of the design process, the results are used as input to design the final robot behavior. We therefore propose to use at this stage an iterative design approach to refine and evaluate the designed behaviors.

4. DISCUSSION

The design process that we proposed integrates technical, artistic and user-centered approaches to design a personality for a robotic application. We started with a user centered perspective to find out what kind of personality people would like a robot to have. Based on this user knowledge, an artistic perspective was taken and ideas for expressions and behaviors of a robot with the particular personality were created. Later in the process, a more technological perspective was taken and the expressions and behaviors were translated into concrete and implementable solutions for a particular robot embodiment (taking into account its requirements and constraints). In the remainder of this section, we summarize the main lessons learned for each of these steps.

4.1 Lessons: Create a personality profile

In order to gather input from users on the desired personality for a domestic robot, we used cards describing personality traits. Upfront, we were uncertain whether people would be able to relate the personality traits to a robot. However, people seemed to have little problems explaining what certain personality characteristics would mean for the behavior of the robot and whether they would appreciate this or not.

Our approach assumes that the application context for the autonomous product is known, because we expect that the task or role of the robot will have a large effect on the personality preference of users. For example, a surveillance robot is expected to have a different personality than a robot that plays games with the user.

Finally, we selected a subset of the Big-Five character traits for our user study that we believed to match the behavior of a domestic robot. We missed, however, a more systematic selection procedure.

4.2 Lessons: Get inspirations for expressions

To get inspiration for expressions we organized a theatre workshop in which actors acted out a domestic robot in various situations. This artistic approach proved useful in inspiring the design of robot behavior. However, we observed that the invited actors who participated in improvisational theatre competitions, were used to express themselves mainly using language and via interactions with the public. Since our main interest is in the expressive movements, we would rather use dancers or mime actors next time. Furthermore, the personality profile restricted the actors in their expression. Of course, the intention of using the personality profile was to guide the actors in their expressions in order to fit the desires of the users. However, it might have limited their creativity.

4.3 Lessons: From actor to robot expressions

The anatomy of the human actors is rather different from the anatomy of the domestic robot we envision. Therefore, it is difficult to map expressions of the actors directly on the robot. However, by abstracting the expression of the actor and keeping the essential characteristics of his movement, we were able to translate it into concrete expressions for the robot.

The sketched storyboard proved to be a fast and useful way to discuss the behavior of the robot. It helped in quickly deciding on a scenario with behaviors to be implemented on the robot. Obviously, the sketches on paper have some limitations. Movements and sounds cannot be realized and require some imagination from the design team.

4.4 Lessons: Visualize in 3D animation

Our main goal for using 3D visualization was to gather qualitative feedback from the user in an early design phase. Animating a virtual version of the domestic robot required less effort than implementing hardware prototypes.

By using physical simulation we gained more realistic behavior that in later stages can be more easily reused on a physical embodiment, but also inherited some the problems of dealing with real world conditions. For example, while in virtual worlds the path of a mobile robot can be repeated exactly in successive runs, physical simulations add some random inaccuracies to the motion, for example due to slip of the wheels. Because successive runs of the same behavior resulted in different output, we chose to show recorded movies of the virtual environment.

Next to these practical experiences, we also want stress some of the more fundamental considerations of this approach that have to be taken into account. Using a virtual simulation of the robot gives the designer the same artistic freedom as in traditional cinematography. The designer has control over the whole scene, including for example lighting, camera angle and other objects in the scene. The camera angle alone can have a significant impact on the perception of the character. In our experiments, we therefore tried to keep the camera in the height of an average person and keep the lighting and objects in the scene as neutral as possible. In reality however, these parameters cannot be controlled, so tests with virtual representations will not substitute for testing the behavior on the physical hardware. This strengthens the argument to create virtual behavior, that can be translated to a physical embodiment.

4.5 Lessons: Evaluate with think-out-loud

From the feedback that we received, we concluded that the participants were able to imagine how the behavior will look like on a physical embodiment, which confirms our assumption that 3D simulations are a good approximation of the physical robot. The qualitative study using a think-out-loud protocol at this early stage of the product development is in our opinion preferred over more quantitative methods. The results give in-depth information about how participants perceived the robot behavior and provide input for redesign of the behavior.

However, the use of a virtual representation of the robot for evaluation has some limitations compared to evaluations with physical robots. For example, simulation of the (physical) interaction between a user and the robot is not possible when using movie clips with animations.

5. CONCLUSION

We have described a process to design the behavior of a domestic robot and proposed it as a way to design a personality and appropriate expressions for autonomous products. The process consists of five main steps, namely creating a personality profile, getting inspiration for the expressions, sketching a scenario, visualizing it in 3D animation, and evaluating it using a think-out-loud protocol. The proposed process combines proven methods from HCI and translates it to the field of HRI. It integrates technical, artistic and user-centered approaches to develop the personality of a robot in an iterative design process. In next steps, we want to improve the process and investigate its applicability in designing a broader range of consumer electronic products. Furthermore, we want to compare our process with existing and widely used product design processes.

Acknowledgements

Our thanks go to Flip van den Berg, Peter Jakobs, Albert van Breemen, Dennis Taapken, and Ingrid Flinsenberg for their valuable contributions to parts of the work described in this paper.

6. REFERENCES

[1] Forlizzi J. and DiSalvo, C. (2006). Service robots in the domestic environment: a study of the roomba vacuum in the home. In: Proceedings HRI ’06, pages 258–265, New York, NY, USA. ACM.

[2] Wrede, B., Haasch, A., Hofemann, N., Hohenner, S., Hüwel, S., Kleinehagenbrock, M., Lang, S., Li, S., Toptsis, I., Fink, G. A., Fritsch, J. and Sagerer, G. (2004). Research issues for designing robot companions: BIRON as a case study. In Proc. IEEE Conf. on Mechatronics & Robotics, volume 4, 1491–1496, Aachen, Germany,

[3] Goodrich, M.A. and Schultz, A.C. (2007). Human-robot interaction: a survey. Found. Trends Hum.-Comput. Interact., 1(3):203–275.

[4] Fong, T., Nourbaksh, I., Dautenhahn, K. (2003). A survey of socially interactive robots, Robotics and Autonomous Systems 42, 143–166

[5] Norman, D. (2001). How humans might interact with Robots, Retrieved Sep. 4th 2008 from http://www.jnd.org/dn.mss/how_might_human.html

[6] Reeves, B. & Nass, C. (1996). The media equation: How people treat computers, televisions, and new media like real people and places, New York: Cambridge University Press

[7] Meerbeek, B., Hoonhout, J., Bingley, P. and Terken, J. (2008). The influence of robot personality on perceived and preferred level of user control. Human and Robot Interactive Communication, Dautenhahn, K. (ed.), 204–229.

[8] Heider, F, & Simmel, M. (1944). An experimental study of apparent behavior. American Journal of Psychology, 57, 243-259.

[9] Tremoulet, P., Feldman, J.(2000). Perception of animacy from the motion of a single object. Perception, volume 29, 943-951

[10] Breazeal, C. (2002). Designing Sociable Robots. MIT Press, Cambridge, MA, USA.

[11] Ishiguro, H. (2006). Interactive humanoids and androids as ideal interfaces for humans. In Proceedings IUI ’06, pages 2–9, New York, NY, USA. ACM Press.

[12] Duffy, B.R. (2003). Anthropomorphism and the social robot. Robotics and Autonomous Systems, 42(3–4):177–190.

[13] McAdams, D., Pals, J. (2006). A New Big Five: Fundamental Principles for an Integrative Science of Personality, American Psychologist vol. 61, nr. 3, 204-217

[14] Carver, C. & Scheier, M. (1995). Perspectives on personality, Boston, USA: Allyn and Bacon

[15] Dryer, D.C. (1999). Getting personal with computers: How to design personalities for agents. Applied Artificial Intelligence, 13(3):273–295.

[16] Kiesler, S.B. & Goetz, J. (2002). Mental models of robotic assistants. CHI Extended Abstracts: 576-577

[17] Walters, M.L., Syrdal, D.S., Dautenhahn, K., Te Boekhorst, R., Koay, K.L. (2008). Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. Autonomous Robot, volume 24 (2)

[18] Duffy, B.R., Dragone, M. and O’Hare, G.M. (2005). The social robot architecture: A framework for explicit social interaction. In Android Science: Towards Social Mechanisms, CogSci 2005 Workshop, Stresa, Italy.

[19] Snibbe, S., Scheeff, M., and Rahardja, K. (1999). A layered architecture for lifelike robotic motion. In Proceedings of The 9th International Conference on Advanced Robotics

[20] Brooks, R.A. (1991). Architectures for Intelligence; VanLehn K. (ed.), chapter How to build complete creatures rather than isolated cognitive simulators, pages 225–239. Lawrence Erlbaum Assosiates, Hillsdale, NJ.

[21] Kawamura, K., Pack, R., and Iskarous, M. (1995). Design philosophy for service robots. In Systems, Man and Cybernetics. Intelligent Systems for the 21st Century, volume 4, pages 3736–3741.

[22] Loyall, A. (1997). Believable Agents: Building Interactive Personalities. PhD thesis, Stanford University.

[23] Neubauer, B.J. (2004). Designing artificial personalities using jungian theory. J. Comput. Small Coll., 20(1):297–305.

[24] Thomas, F. and Johnson, O. (1981). The Illusion of Life - Disney Animation. Walt Disney productions.

[25] Van Breemen, A. (2004). Bringing robots to life: Applying principles of animation to robots. In Proceedings of the Workshop on Shaping Human-Robot Interaction - Understanding the Social Aspects of Intelligent Robotic Products (CHI2004), Vienna.

[26] Dautenhahn, K. (2002). Design spaces and niche spaces of believable social robots. In Robot and Human Interactive Communication, pages 192–197.

[27] Benyon, D., Turner, P. and Turner, S. (2004). Designing interactive systems. Addison-Wesley New York.

[28] Gould, J.D. and Lewis, C. (1985). Designing for usability: key principles and what designers think. Commun. ACM, 28(3):300–311.

[29] Cockton, G. (2008). Revisiting usability’s three key principles. In CHI ’08: CHI ’08 extended abstracts on Human factors in computing systems, pages 2473–2484, New York, NY, USA. ACM.

[30] Ljungblad, S., Walter, K., Jacobsson, M. and Holmquist, L. (2006). Designing personal embodied agents with personas. In Robot and Human Interactive Communication, ROMAN 2006., pages 575 – 580.

[31] Pruitt, J. and Grudin, J. (2003). Personas: Practice and theory. In Proceedings of the 2003 Conference on Desiging for User Experiences.

[32] Osada, J., Ohnaka, S. and Sato, M. (2006). The scenario and design process of childcare robot, papero. In proceedings ACE ’06, page 80, New York, NY, USA. ACM.

[33] Friess, E. (2008). Defending design decisions with usability evidence: a case study. In CHI ’08: extended abstracts on Human factors in computing systems, pages 2009–2016, New York, NY, USA. ACM.

[34] Höök, K. (2004). From Browse to Trust: evaluating embodied conversational agents, volume 7 of Human-Computer Interaction Series, chapter User-Centred Design

and Evaluation of Affective Interfaces: A Two-tiered Model, pages 127–160. Springer Netherlands.

[35] Srivastava, S. (2008). *Measuring the Big Five Personality Factors.* Retrieved Sep.4th 2008 from http://www.uoregon.edu/~sanjay/bigfive.html.

[36] John, O.P. et al., The Big-Five Trait Taxonomy: History, Measurement, and Theorethical Perspectives, In: Pervin, L. and O.P. John, Handbook of personality: Theory and research, New York, 1999

[37] Costa, P. and R. McCrae (2008). NEO Personality Inventory – Revised, Retrieved Sep. 4th 2008 from http://www3.parinc.com/products/product.aspx?Productid=NEO-PI-R

[38] Boeree, G. (2004). Big Five Mini Test. Retrieved Sep. 4th 2008 from http://www.ship.edu/~cgboeree/bigfiveminitest.html

[39] Saucier, G. (1994). Mini-Markers: A Brief Version of Goldberg's Unipolar Big-Five Markers, Journal of Personality Assessment, 63(3), 506-516

[40] Stanislavski, C. and Hapgood, E. R. T. (1889). Building a Character. Routledge.

[41] Trappl, R., Petta, P. and Payr, S. editors, 2003. Emotions in Humans and Artifacts. The MIT Press.

[42] Strack, F., Martin, L. and Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. Journal of Personality and Social Psychology, 54(5):768–777.

[43] Van Loenen, E., De Ruyter, B. & Teeven, V. (2006). ExperienceLab: Facilities. In Aarts, E. & Diederiks, E. (Eds.), Ambient Lifestyle: From Concept to Experience (pp. 47 – 53). Amsterdam, The Netherlands: BIS Publishers

[44] Buur, J. & Søndergaard, A. (2000). Video card game: an augmented environment for user centred design discussions. Designing Augmented Reality Environments 2000: 63-69

[45] Ruttkay, Z., Hendrix, J., Ten Hagen, P., Leli`evre, A., Noot, H. and De Ruyter, B.(2000). A facial repertoire for avatars. In Proceedings of CELE-Twente Workshops on Natural Language Technology, pages 27–46.

[46] Saerbeck, M. and Van Breemen, A.J. (2007). Design guidelines and tools for creating believable motion for personal robots. In Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on, pages 386–391.

[47] Kidd, C. and Breazeal, C. (2004). Effect of a robot on user perceptions. In Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on, volume 4, pages 3559 – 3564. IEEE.

[48] Powers, A., Kiesler, S., Fussell, S. and Torrey, C. (2007). Comparing a computer agent with a humanoid robot. In HRI '07: Proceeding of the ACM/IEEE international conference on Human-robot interaction, pages 145–152, New York, NY, USA. ACM Press.

[49] Van Breemen, A.J. (2005). Scripting technology and dynamic script generation for personal robot platforms. In Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on, pages 3487 – 3492. IEEE CNF.

Affective-Centered Design for Interactive Robots

Laurel D. Riek[1] and **Peter Robinson**

Abstract. We present a new paradigm for the design of interactive robots called affective-centered design. By drawing on the disciplines of human-computer interaction (HCI), affective computing, and human-robot interaction (HRI), we suggest techniques robot designers can use to help ensure interactions with their robots are of high affective quality, and thus more likely to be enjoyed and accepted by users.

Category: Position Paper (*P*)

1 INTRODUCTION

Each year, robots are entering domestic environments in greater and greater numbers. According to a 2007 report by the International Federation of Robotics, there were 3.4 million personal service robots in use worldwide in domestic settings. The report forecasts that this number is expected to increase by 4.6 million robots by 2012 [40]. These domestic robots are being used to serve as health aids and companions, help with household chores, and provide education and entertainment to their users.

The domestic robot user presents a unique challenge to robot designers. Elderly users are likely to be uncomfortable with domestic robots due to a lack of exposure to technology, disabled users might have difficulty using robots that do not provide interaction modalities that accommodate their needs, and people using robots for household chore assistance are unlikely to have much time to devote to learning to use complexly designed systems. One way to address some of these problems is to design robots that allow people the ability to interact with robots naturally.

Natural interaction means allowing people the ability to communicate with robots in ways similar to how they communicate with other people. This includes both verbal communication (speech and non-speech vocalization) and nonverbal communication (body gesture, gaze, movement, and facial expression). Most people are able to express themselves in this way and easily interpret such expressions in others. While people generally do not expect such ease of interaction with machines, evidence suggests having it would help improve user engagement with the robot [57]. Indeed, by taking advantage of these interactive modalities, robot designers can go a long way toward ensuring their robots are accepted.

Thus, we present a new paradigm for interactive robot design called affective-centered design, which we describe in Section 2. This paradigm draws on the fields of human-computer interaction (HCI), human-robot interaction (HRI), and affective computing. Our desired goal for presenting this paradigm is to help designers create a positive user experience by ensuring high affective quality of interaction with a robot.

[1] Computer Laboratory, University of Cambridge, United Kingdom, email: Laurel.Riek@cl.cam.ac.uk

1.1 Affective quality

Before describing the concept of affective quality and how it relates to interactive robot design, it is important to explicitly define the concept of affect. Many researchers struggle with precisely defining words like “affect”, “emotion” , and “mood”, and rightfully so. The definitions of these words have become increasingly convoluted not only in vernacular English, but within scholarly literature as well. For the purposes of this paper, we shall adopt Clore and Palmer’s stance which is to simply include these concept under the umbrella term “affective states.” [2]. Thus, “affect” refers to “an embodied reaction of pleasure or displeasure signifying the goodness or badness of something.’ A “psychological ‘state’ is assumed to exist whenever multiple systems of an organism reflect the same condition at the same time.” For example, anger as an affective state might refer to both internal thoughts and feelings, as well as non-verbal expressions of anger (furrowed eyebrows, a raised voice, etc.) [10].

Affective quality is the “ability of an object or stimulus to cause changes in one’s [affective state]” [67]. To call a human’s interaction with a robot to be of “high affective quality” means that the overall affective state of the user is positive during the interaction. This is a qualitative yet objective assessment of interaction quality. As we will see in Section 4.1, HCI researchers have successfully used affective quality as a usability dimension when assessing systems. Thus, we believe it will also be useful as a usability dimension for interactive robots.

1.2 Implicit communication

Palen and Bødker [41] suggest that emotions should not be seen as simply a *feature* of interaction, but as an integral part of interaction itself. Thus, all interaction is emotional; even when it is emotionally neutral it is still framed by the idea of emotions. In some ways, emotions can be seen as a communication medium. Rani and Sarkar [43] describe this as “implicit communication”, where a robot responds to the emotional state of a human (e.g., afraid, tired, happy) or simply to an emotionally neutral intention of a human (e.g., *Fetch me that object I’m pointing at*).

Implicit communication is an important line of research for interactive robot designers, because the ability for robots to fully understand explicit human communication (i.e., speech) is a long way off. While the natural-language processing community continues to advance the state-of-the-art in this area, roboticists can make huge strides forward by creating robots that understand at least rudimentary non-verbal behavior.

Communication is a two-way street, however. As Wallach and Allen point out, robots that interact with people in a social context

[2] Within this paper, if we use the words “emotion”, “affect”, or “mood”, we always mean the umbrella term “affective states” as defined in Section 1.1.

should be able to gesture in a way that allows them to communicate their intentions [64]. This allows people the ability to understand a robot's capabilities and know what to expect. Cañamero and Gaussier [9] describe this in terms of a dichotomy between internal and external communication. Internal being the affective states the robot represents (including its representation of the emotional context outside itself) and external being the outward manifestation of emotion that the robot makes.

Thus, affective-centered design for interactive robots has two important aspects. There is affect recognition - *what is the human's affective state*? And affect generation - *how should the robot respond to that state and express itself appropriately?*

2 AFFECTIVE-CENTERED DESIGN

Affective-centered design is an iterative design process modeled somewhat after human-centered design. Human-centered design is an ISO standard that gives guidance throughout the entire design lifecycle of an interactive system. (See Figure 1). Some of its major tenets include:

- Involving end-users throughout the design lifecycle to ensure their needs are adequately addressed
- Understanding the context in which the system will be used
- Appropriately delegating function between the system and the user
- Adequately understanding the capabilities of both the system and the user
- Iterative design
- Multi-disciplinary design[53]

While this standard provides a general framework for interactive robot design, it does not provide practical guidelines that robot designers can employ. Furthermore, it does not consider affective quality whatsoever within the design process.

Therefore, in addition to the aforementioned tenets, the affective-centered design process also includes the following tenets, all of which can be employed iteratively throughout the design lifecycle of an interactive robot:

- Using affective quality as a metric throughout the design lifecycle
- Surveying the affective states of users
- Evaluating the affect generation of robots

We will now explain each of these tenets in detail. Later, in Section 3, we will describe a real-world example of how we employed the affective-centered design process in our own research. (Also, see Figure 2 for an illustration of the process.)

2.1 Affective quality throughout the entire lifecycle

In HCI, affective quality has been used successfully as an evaluative metric for interactive systems (See Section 4.1). And in HRI, several robots have been designed with user and/or robot affect in mind (See Section 4.3). However, a number of these projects examined human responses very late in the design lifecycle, often well after the systems had already been built. For 2D interactive systems this is somewhat acceptable, because it is relatively easy to change graphical-user interfaces. However, for robots, physical hardware changes are far more problematic. For example, if a humanoid robot is built with an extremely frightening face, no amount of behavioral changes to its affect generation software will change the fact that its physical

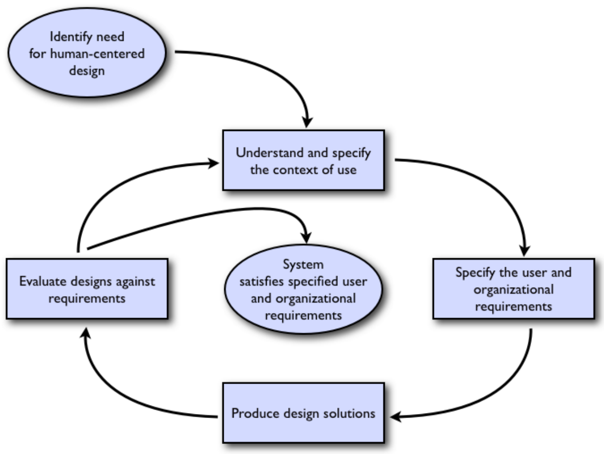


Figure 1. The human-centered design process for interactive systems. [23]

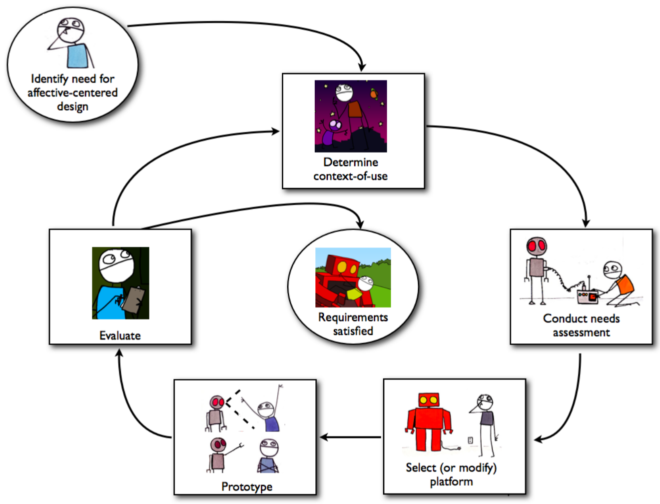


Figure 2. The affective-centered design process for interactive robots. Images by Sam Brown, explodingdog.

appearance is scary. Therefore, it is critical that interactive robot designers begin to present design ideas to users before a single wire is soldered, and continue to do so throughout the entire design lifecycle of the robot.

It is possible to present design ideas to users in an inexpensive way via paper prototypes, photographs, and simulation software. This need not be a formal usability study; it can be accomplished by talking to friends, colleagues, and family members. It is best to include at least some users from the population of people intended to interact with the robot, but this is not always possible. The most important thing is that representative users are included. For example, if one is designing a companion robot for the elderly, presenting design ideas to one's great aunt will yield far more informative results than solely presenting ideas to one's labmate.

If during the design process it becomes apparent that the affective quality of interacting with the robot is negative, it is important to identify why and figure out ways to mitigate it. Users are generally able to explicitly describe what they dislike about a robot, though it may require some in-depth interviewing to isolate the problem.

For example, we recently designed a empathetic conversation robot (described in Section 3) and ran a pilot study to evaluate it. After our study, we asked subjects how they felt about conversing with the robot. Here is an excerpt of one interview:

Experimenter: How did you feel talking to the robot?
Subject: Very weird. I've done voice recognition training, so I've talked to computers before. But [the robot] wasn't trying to make me feel anything. It was a weird experience.
Experimenter: How was it weird?
Subject: The reactions he had to what I was saying seemed to be negative.
Experimenter: How so?
Subject: He looked away from me at the start of the experiment. I would expect him to nod as if he was interested [in what I was saying], not look away.

While a user might give a terse reply, "It was a weird experience," by careful follow-up questioning it can quickly become clear what the problem is (in this case, poor gaze direction).

2.2 Surveying users' affective states

There are two primary ways to survey users' affective states during the interactive robot design lifecycle. The first is via a human experimenter who is in some way assessing a user. Such an assessment might be made using ethnographic observation [18], common ground analysis [58], embodiment analysis [12], questionnaires, interviews, or other techniques [19]. The goal of these techniques is to characterize the interaction with the robot in terms of people's attitudes toward it.

The second way is to perform automatic evaluation of users' affective states. This type of assessment is most likely to be performed by the robot itself, and will be used to modify its own behavior accordingly. We will describe this in more detail in Section 4.2.

Both of these affective state analyses are useful throughout the design lifecycle of an interactive robot. By understanding users' attitudes one can begin to characterize the quality of interaction, and modify the robot's design as necessary.

2.3 Evaluating affect generation of robots

The third tenet of affective-centered design concerns evaluating how the robot expresses affect. This measure is, of course, tightly coupled with how the user feels about the interaction. But this might also be something evaluated by robot designers independent of a user's reactions. For example, during a prototyping stage, certain behaviors might need be tweaked in order to produce the desired result.

A robot's affective state can be assessed on several levels, such as its contextual understanding of the world, the appropriateness of responses it gives, and the degree to which it expresses an affective state. To perform these assessments, robot designers can actually employ some of the same techniques as described in the Section 2.2, particularly ethnography and common-ground analysis. The difference is that the robot becomes the focus of evaluation, and its affective expression is what is measured.

3 AFFECTIVE-CENTERED DESIGN IN PRACTICE

To illustrate how robot designers can use affective-centered design in practice, we will briefly present our ongoing work on a conversational robot. In our work with this robot we have employed several elements of affective-centered design, such as understanding the context of use, carefully selecting our platform, performing design tradeoffs, running a pilot study, and iterating on Virgil's design accordingly.

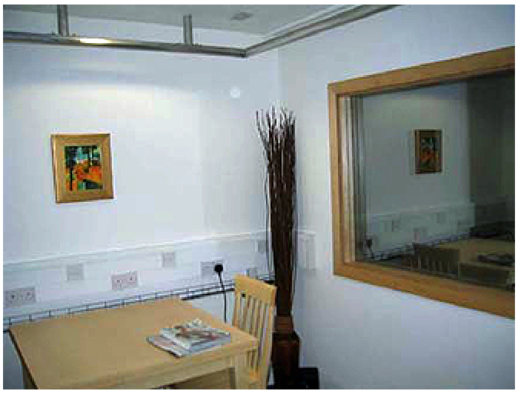


Figure 3. The Computer Laboratory usability lab.

3.1 Context of use

Before we began our research, we first considered the context of use for our robot. We were interested in creating an empathetic conversation partner capable of understanding some rudimentary affective states of humans and responding to them with appropriate affective expressions of its own.

Since the focus of our research is on empathy during one-on-one human conversations, we decided that the most appropriate physical space for users to interact with the robot would be in a home-like setting. Thus, we selected the Cambridge Computer Laboratory usability lab as the place to perform our initial studies. The usability lab is meant to resemble a living room - it has carpets, tables, chairs, and pictures on the wall. It also has a one-way mirror to allow for unobtrusive observation (See Figure 3).

3.2 Platform selection

While we ultimately would like to use an expressive humanoid head for our research, at present we are limited to inexpensive, off-the-shelf robots on which to try our ideas. The robot we selected is actually in some ways a prototype; by only costing \$150 USD we are able to make initial strides forward in the behavioral design of an empathetic robot.

Thus we chose to begin our work on the robot Virgil, a chimpanzee robot made by WowWee. (See Figure 4). Virgil is a robot with 18 degrees-of-freedom (DOFs) in total. Its upper-lip has 2 DOFs (up/down), its lower jaw 2 DOFs (up/down), its eyes 4 DOFs (up/down/left/right), eyebrows 2 DOFs (up/dpwn), and its head 8 DOFs (roll/pitch/yaw). The robot can be operated via remote control and it also has some autonomous reactive behaviors, such as moving its head, opening and closing its mouth, and scrunching its nose [44].

3.3 Design tradeoff analysis

Since Virgil is intended to be a conversational robot, we initially wanted to have the robot speak. However, as we began working with

Figure 4. Virgil, our conversation robot.

the robot we quickly realized that performing speech-to-mouth syntonization is extremely difficult, and with a robot with only 4 DOFs in its lower face we realized it would be quite challenging to do a convincing job. As is well-documented in the psychology and animation literature, people are so sensitive to incongruous speech-face synchronization it can interfere with speech comprehension [33].

Thus, we decided it was better to have a silent robot than one that made people uncomfortable due to its strange mouth movements. Ironically this design tradeoff actually ended up lowering the affective quality of the robot in a way we had not predicted, as we will see in Section 3.4.

3.4 Pilot study

Early on in our design lifecycle, we wanted to explore the ideas of how non-verbal communicative gestures of empathy might alter people's affective states. So we ran a pilot study [44] with Virgil and a user interacting one-on-one in the usability lab. We asked all subjects to tell both a non-personal and personal story to the robot.

Our study had an experimental and control condition. In the test condition we wizard-of-oz controlled the robot to make empathetic head gestures (nodding) and mouth movements (open mouth in surprise). In the control condition, the robot acted autonomously in a random manner.

Following the study, we asked subjects to complete a written questionnaire that asked them to rate their interaction with the robot. These were 5-point Likert scale questions such as "I think Virgil could be a friend of mine," and "Virgil recognized my feelings and emotions appropriately for the situation". This resulted in each subject having an overall interaction satisfaction score. As we predicted, subjects in the experimental condition rated the interaction far more favorably than those in the testing condition [44].

We also asked subjects follow-up interview questions such as, "How did you feel talking to the robot?" and "Did you feel the robot was an amicable conversation partner?". Subjects made comments about the appearance of Virgil (ranging from neutral to negative), the appropriateness of responses it gave (people in the control condition made very negative remarks, while those in the experimental group were more positive), and communication flow. Most surprisingly, nearly every subject made strong statements that they wished the robot would make communicative noises or else spoke back. They said conversation was a two-way street, and one-sided conversations just did not feel natural to them [44].

This study was very helpful to us because it made us rethink several major design decisions, including our decision to have the robot be silent. Since we are in the early stages of our work we are fortunately able to modify the platform to accommodate these insights. We will then perform an additional pilot study to see if these design changes lead to an improved quality of interaction.

4 RELATED WORK

Since the goal of affective-centered design is to improve the affective quality of interaction with a robot, there are (at least) three primary fields that help to inform this design process. They are: human-computer interaction (HCI), affective computing, and human-robot interaction (HRI). This section provides a brief overview of relevant work done in these fields as they relate to affective-centered design for interactive robots. By no means is this section exhaustive, nor is it the case that these fields are mutually exclusive.

4.1 Human-computer interaction

In the field of HCI, a transition has begun to take place within the community from "human factors to human actors" [1]. In other words, researchers are considering people's emotional experience while interacting with a system as a new dimension of usability. Empirical research has shown that how people perceive the affective quality of a system positively impacts how they perceive its usability [11, 50, 59, 60, 67].

Norman has done pioneering work on the idea of using affective quality as a usability dimension. He suggests that when people "feel good" about a system they are more likely to overlook flaws in its design, find the system easier to learn, and also perform better when confronted with difficult tasks [36]. Later, Norman proposed a theoretical framework for three levels of emotional experience: Reflective, Behavioral, and Visceral.

- The reflective layer concerns "intellectually-induced reactions". So, in the example of product design, a Perrier label on a water bottle.
- The behavioral layer concerns "expectation-induced reactions". For example, the water inside a plastic bottle.
- The visceral layer concerns "perceptually-induced reactions". This would be, "a beautiful bottle that is used as a vase." [37, 38]

This framework has received recent empirical support by Lim et al. who found that the reflective layer was tightly coupled to peoples' experience of usefulness, as were the other layers. Thus, it is important for interactive products to have useful functionality. Furthermore, the researchers found that the overall quality of the interaction is critical to people's emotional experience [30].

This result has interesting design implications for interactive robots, because a robot's appearance might not be in concordance with its function or behavior [45]. For example, it may be difficult to initially identify a companion robot as a companion, whereas it is usually pretty easy to identify a telephone as a telephone. Thus, novel, interactive robots may need to present affective clues to users to appropriately advertise their function.

4.2 Affective computing

Affective computing is a discipline devoted to the idea of giving machines the ability to recognize and generate affect [42]. In some ways, the field exists to address the failings of traditional HCI systems, which typically neglect affective state changes in users. In fact, some argue that such neglect is a reason many users view interactions with computers as "cold, incompetent and socially inept." In order to address this, several leaders in the field have stated that it is critical that user interfaces of the future are able to "detect subtleties of and changes in the user's behavior, especially his/her affective behavior, and to initiate interactions based on this information rather than simply responding to the user's commands" [66].

Until recently, most of the approaches to affect recognition centered around posed data with exaggerated affective expressions, were limited to a small set of emotions (such as anger, fear, and happiness), and were restricted to single modes of expression (just face or just speech). However, the field is now shifting toward looking at recognizing, multi-modal, less-constrained naturalistic expressions [66]. For example, el Kaliouby and Robinson worked on the generalization of facial affect inference for complex mental states [14, 15, 16] while Sobol and Robinson worked on inferring affect from naturally-evoked speech [55, 56]. Bernhardt and Robinson worked on inferring affect from body posture and gesture [2].

As for affect generation (also called "emotion synthesis"), research into several affective channels of expression are being explored. These include facial animation [13], gestures [20], speech [51], nonverbal vocalizations, and others.

The field of affective computing is very relevant to interactive robot designers because the community has already begun to tackle a number of hard problems related to interacting in the physical world. Poor lighting, noisy environments, sensor fusion, widely varying communication styles, and other problems are also encountered in robotics.

4.3 Human-robot interaction

In the field of HRI, quite a number of interactive robots have been designed with affect in mind. Breazeal et al. [4] and Fong et al. [17] present thorough surveys of many such robots and their theoretical emotional underpinings. We will present a subset of these robots and introduce a few others using role categories suggested by Scholtz [52] and Goodrich and Schultz [19], as well as some general topic-area categories we've created. For each category we will list the names of representative robots and cite papers about them that give mention to the affective aspects of their design.[3] We will then highlight one robot and discuss how its design incorporates elements of affective-centered design. (The name of the highlighted robot will be italicized.)

4.3.1 Epigenetic (Developmental) Robots

Robots: Cog[49], HOAP-3[8], iCub[62], *Kismet[3]*, Leonardo[32]

A number of interactive robots have been created with some degree of affective understanding and generation capability using an epigenetic approach. This approach uses ideas from developmental psychology to help robots learn sophisticated social behaviors [48]. Many of these developmentally-based robots inherently take social context into account in order to learn to adapt to the humans interacting with them. One of the first of these robots is Kismet, an anthropomorphic, expressive robot designed entirely for emotional interaction with humans. By understanding the social cues of humans in the environment, Kismet is able to respond in an emotionally appropriate way to people [3]. Its thoughtful design has lead to it being a very well accepted and regarded robot.

[3] This listing is by no means exhaustive, nor are these categories mutually exclusive.

4.3.2 Entertainment Robots

Robots: AIBO[24], ASIMO[35], *Keepon*[26], AUR[22], Improv Robots[6]

Kozima and Michalowski were interested in building a robot that could interact with children in a pleasant and natural way. Their first attempt was the Infanoid robot, which was a highly mechanical-looking, very expressive robot. From observational studies the researchers found that the appearance and behavior of this robot was overwhelming children. This insight led them to the successful design of the robot Keepon, which is a minimally-designed interactive dancing robot. The robot only has 4 degrees of freedom, but is easily able to express attention via head direction and emotion via rocking motions. Its design was well informed by observing hundreds of children interacting with the robot for over 400 hours in total [26].

4.3.3 Therapeutic Robots

Robots: Huggable[29], iCat[21], KASPAR[46], *PARO*[63], Shybot[28],

Shibata et al. describe their desire to build an affect robotic pet that was capable of sensing the emotions of the people it was interacting with and alter its affect accordingly [54]. Their design description indicates an implicit understanding of affective-centered design, because from the outset they concerned themselves with how their robot would interact emotionally with users, and tailored the robot's design accordingly. This mindset led the reearchers to later create the very successful implementation of PARO the robotic seal, which has been used effectively to reduce stress and depression among the elderly [63].

4.3.4 Peer Robots

Robots: Vikia[7], Robonaut[61], *Valerie* [25], GRACE[47], Mel [57]

Kirby et al. designed Valerie, a robot receptionist designed to facilitate long-term social interaction with people; the researchers wanted the robot to maintain people's interest over time [25]. The robot was thoughtfully designed in a way that employed elements of affective-centered design - the physical appearance of the robot, its station, and its behaviors were carefully considered to create an engaging experience with users. After several years of observations of people interacting with the robot, the researchers also realized ways to improve its design that were motivated by the emotional state of users [47].

4.3.5 Mentor Robots

Robots: *Basketball Coach* [31], Chips[39], RoCo[5]

Liu et al. describe a robotic basketball coach that monitored the physiological signals (heartrate and galvanic skin response) of people while they shot baskets. Depending on how anxious people seemed to be, the robot altered the game's level of difficulty. The researchers found through this style of interactive teaching people's performance improved [31]. This design is very much affective-centered within the context of a closed loop system. However, it's worth noting that requiring people to wear sensors is often undesirable due to problems with affixing sensors to the skin via abrasive gels, calibration problems, and the hardware being seen as obtrusive [65]. Therefore, having users endure such issues might not outweigh the benefit of having robots "in the know" about people's physiology. This sort of decision making is part of the tradeoff analysis phase that needs to be considered during the affective-centered design process.

4.3.6 Industrial Robots

Robots: *Safety Arm*[27], WE-4R Arm [34]

One normally would not expect designers of industrial robots to need to consider affect in their design; however, those that work near humans may very well need to be aware of their surrounding social context. For example, Kulic and Croft describe using affect recognition in an interactive scenario where an industrial robotic arm and a human are working together. The robot is made aware of the user's affective state (in response to its motions), and uses this information to calculate a "danger index", thus modifying its behavior accordingly [27]. This sort of thinking is also employing elements of affective-centered design - the researchers knew that industrial robots working with humans involve safety risks, and it would be helpful if such robots could quickly understand when their human peers feel afraid so that they do not endanger them.

The aforementioned work in HRI helps illustrate the value of using affective-centered design practices when designing interactive robots. By considering the affective states of users as well as the affect generation capabilities of robots, robot designers are helping to ensure their robots will be well-accepted and understood by people.

5 DISCUSSION

We introduced affective-centered design, which is a new process for the design of interactive robots. The motivation behind such an process is to give robot designers techniques for ensuring that interaction with their robots is of high affective quality, meaning that the overall affective state of the user is positive during the interaction. This will hopefully help ensure more people are accepting of robots in domestic environments.

The major tenets of affective-centered design are very similar to human-centered design, where one takes time to understand the robot's context of use by involving representative users throughout the design lifecycle and performing iterative, multi-disciplinary design. This is accomplished by using affective quality as an evaluative measure, through both surveying the affective states of users and evaluating the affect generation capability of robots. We demonstrated how one might go about this process in practice by discussing how we developed the conversational robot Virgil.

Affective-centered design is a process that sits at the intersection of three fields - HCI, affective computing, and HRI. These fields are all interested in ensuring technology is well accepted by end-users; thus, by examining affect as a quality of interaction with robots, we hope the affective-centered design process will prove helpful to researchers in each of these fields.

ACKNOWLEDGEMENTS

This work is supported by the QUALCOMM Research Studentship. We would also like to thank Shazia Afzal for her assistance.

REFERENCES

[1] L. Bannon, 'From human factors to human actors: the role of psychology and human-computer interaction studies in system design', in *Design at work: cooperative design of computer systems*, eds., J. Greenbaum and M. Kyng, 25–44, L. Erlbaum Associates Inc., Hillsdale, NJ, USA, (1992).

[2] D. Bernhardt and P. Robinson, 'Detecting affect from non-stylised body motions', in *Proc. of the 2nd international conference on Affective Computing and Intelligent Interaction (ACII)*. Springer-Verlag, (2007).

[3] C. Breazeal, *Designing Sociable Robots*, MIT Press, Cambridge, MA, USA, 2002.

[4] C. Breazeal, A. Takanishi, and T. Kobayashi, 'Social robots that interact with people', in *Springer Handbook of Robotics*, eds., B. Siciliano and O. Khatib, 1349–1369, Springer, (2008).

[5] C. Breazeal, A. Wang, and R. Picard, 'Experiments with a robotic computer: body, affect and cognition interactions', in *Proc. of the ACM/IEEE international conference on Human-robot interaction (HRI)*, pp. 153–160, Arlington, Virginia, USA, (2007). ACM.

[6] A. Bruce, J. Knight, S. Listopad, B. Magerko, and I. R. Nourbakhsh, 'Robot improv: using drama to create believable agents', in *Proc. International Conference on Robotics and Automation (ICRA)*, (2000).

[7] A. Bruce, I. Nourbakhsh, and R. Simmons, 'The role of expressiveness and attention in human-robot interaction', in *Proc. IEEE International Conference onRobotics and Automation (ICRA)*, (2002).

[8] S. Calinon and A. Billard, 'A framework integrating statistical and social cues to teach a humanoid robot new skills', in *Proc. IEEE Intl Conf. on Robotics and Automation (ICRA), Workshop on Social Interaction with Intelligent Indoor Robots*, (May 2008).

[9] L. Canamero and P. Gaussier, 'Emotion understanding: Robots as tools and models', in *Emotional development: Recent research advances*, eds., J. Nadel and D. Muir, 235–258, Oxford University Press, New York, NY, (2005).

[10] G. L. Clore and J. Palmer, 'Affective guidance of intelligent agents: How emotion controls cognition', *Cognitive Systems Research*, **10**(1), 21 – 30, (2009).

[11] A. Crystal and S. Kalyanaraman, 'Usability, cognition, and affect in web interfaces: The role of informative feedback and descriptive labeling', in *International Communication Association*, (2004).

[12] K. Dautenhahn, B. Ogden, and T. Quick, 'From embodied to socially embedded agents - implications for interaction-aware robots', *Cognitive Systems Research, Special issue on Situated and Embodied Cognition*, **3**(3), 397–428, (2002).

[13] Z. Deng and J. Y. Noh, 'Computer facial animation: A survey', in *Data-Driven 3D Facial Animation*, Springer, (2007).

[14] R. el Kaliouby, *Mind-reading Machines: the automated inference of complex mental states from video*, Ph.D. dissertation, Computer Laboratory, University of Cambridge, 2005.

[15] R. el Kaliouby and P. Robinson, 'Generalization of a vision-based computational model of mind-reading.', in *In Proc. International Conference on Affective Computing and Intelligent Interaction*, eds., J. Tao, T. Tan, and R. W. Picard, volume 3784 of *Lecture Notes in Computer Science*, pp. 582–589. Springer, (2005).

[16] R. el Kaliouby and P. Robinson, 'Real-time inference of complex mental states from facial expressions and head gestures', in *Real-time Vision for Human-Computer Interaction*, 181–200, Springer, (2005).

[17] T. Fong, I. Nourbakhsh, and K. Dautenhahn, 'A survey of socially interactive robots', *Robotics and Autonomous Systems*, **42**, 143–166, (March 2003).

[18] J. Forlizzi, 'How robotic products become social products: an ethnographic study of cleaning in the home', in *Proc. of the ACM/IEEE international conference on Human-robot interaction*, pp. 129–136, New York, NY, USA, (2007). ACM.

[19] M. A. Goodrich and A. C. Schultz, 'Human-robot interaction: a survey', *Foundations and Trends Human-Compututer Interactaction*, **1**(3), 203–275, (2007).

[20] B. Hartmann, M. Mancini, and C. Pelachaud, 'Implementing expressive gesture synthesis for embodied conversational agents', *Gesture in Human-Computer Interaction and Simulation*, 188–199, (2006).

[21] M. Heerink, B. Krose, B. Wielinga, and V. Evers, 'Enjoyment intention to use and actual use of a conversational robot by elderly people', in *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction (HRI 08)*, pp. 113–120, New York, NY, USA, (2008). ACM.

[22] G. Hoffman, R. Kubat, and C. Breazeal, 'A hybrid control system for puppeteering a live robotic stage actor', in *The 17th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, (Aug. 2008).

[23] ISO/IEC, '13407 Human-centred design processes for interactive systems', Technical report, ISO/IEC, (1999).

[24] P.H. Kahn Jr., B. Friedman, D. R. Perez-Granados, and N. G. Freier, 'Robotic pets in the lives of preschool children', in *CHI '04 extended abstracts on Human factors in computing systems*, pp. 1449–1452, New York, NY, USA, (2004). ACM.

[25] R. Kirby, F. Broz, J. Forlizzi, M. P. Michalowski, A. Mundell, S. Rosenthal, B. P. Sellner, R. Simmons, K. Snipes, A. Schultz, and J. Wang, 'Designing robots for long-term social interaction', in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2199 – 2204. IEEE, (August 2005).

[26] H. Kozima, M. Michalowski, and C. Nakagawa, 'Keepon: A playful robot for research, therapy, and entertainment', *International Journal of Social Robotics*, (2008).

[27] D. Kulic and E. A. Croft, 'Affective state estimation for humanrobot interaction', *Robotics, IEEE Transactions on*, **23**(5), 991–1000, (2007).

[28] C. J. Lee, K. Kim, C. Breazeal, and R. Picard, 'Shybot: friend-stranger interaction for children living with autism', in *CHI '08 extended abstracts on Human factors in computing systems*, pp. 3375–3380, Florence, Italy, (2008). ACM.

[29] J. K. Lee, R. L. Toscano, W. D. Stiehl, and C. Breazeal, 'The design of a semi-autonomous robot avatar for family communication and education', in *The 17th International Symposium on Robot and Human Interactive Communication (RO-MAN 08)*, pp. 166–173. IEEE, (2008).

[30] Y. Lim, J. Donaldson, H. Jung, B. Kunz, D. Royer, S. Ramalingam, S. Thirumaran, and E. Stolterman, 'Emotional experience and interaction design', in *Affect and Emotion in Human-Computer Interaction: From Theory to Applications*, eds., C. Peter and R. Beale, 116–129, Springer-Verlag, Berlin, Heidelberg, (2008).

[31] C. Liu, P. Rani, and N. Sarkar, 'Human-robot interaction using affective cues', in *Proc. of the 15th International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, (September 2006).

[32] A. Lockerd and C. Breazeal, 'Tutelage and socially guided robot learning', *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems, (IROS).*, **4**, 3475–3480 vol.4, (2004).

[33] H. Mcgurk and J. Macdonald, 'Hearing lips and seeing voices', *Nature*, **264**(5588), 746–748, (December 1976).

[34] H. Miwa, K. Itoh, D. Ito, H. Takanobu, and A. Takanishi, 'Design and control of 9-dofs emotion expression humanoid arm', in *Proceedings of the 2004 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 128–133. IEEE, (2004).

[35] B. Mutlu, J. Forlizzi, and J. Hodgins, 'A storytelling robot: Modeling and evaluation of human-like gaze behavior', in *Proc. International Conference on Humanoid Robots*. IEEE, (December 2006).

[36] D. A. Norman, 'Emotion & design: attractive things work better', *Interactions*, **9**(4), 36–42, (2002).

[37] D. A. Norman, *Emotional Design: Why We Love (or Hate) Everyday Things*, Basic Books, May 2005.

[38] D. A. Norman, A. Ortony, and D. M. Russell, 'Affect and machine design: Lessons for the development of autonomous machines', *IBM Syst. J.*, **42**(1), 38–44, (2003).

[39] I.R. Nourbakhsh, C. Kunz, and T. Willeke, 'The mobot museum robot installations: a five year experiment', in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, (2003).

[40] International Federation of Robotics, 'World robotics 2008: Statistics, market analysis, forecasts, case studies, and profitability of robot investment', Technical report, International Federation of Robotics Statistical Department, (October 2008).

[41] L. Palen and S. Bødker, 'Don't get emotional', in *Affect and Emotion in Human-Computer Interaction: From Theory to Applications*, Lecture Notes in Computer Science (LNCS), 12–22, Springer-Verlag, Berlin, Heidelberg, (2008).

[42] R. Picard, *Affective Computing*, MIT Press, 1997.

[43] P. Rani and N. Sarkar, 'Emotion-sensitive robots - a new paradigm for human-robot interaction', in *Proc. of the IEEE-RAS/RSJ International Conference on Humanoid Robots (Humanoids 04)*, pp. 10–12, (2004).

[44] L. D. Riek and P. Robinson, 'Real-time empathy: Facial mimicry on a robot', in *Workshop on Affective Interaction in Natural Environments (AFFINE) at the International ACM Conference on Multimodal Interfaces (ICMI 08)*. ACM, (2008).

[45] L. D. Riek and P. Robinson, 'Robot, rabbit, or red herring? Societal acceptance as a function of classification ease', in *17th International IEEE Symposium on Robot and Human Interactive Communication (RO-MAN 2008) Workshop on Robots as Social Actors*. IEEE, (2008).

[46] B Robins, K. Dautenhahn, R te Boekhorst, and C. L. Nehaniv, 'Behaviour delay and robot expressiveness in child-robot interactions: a user study on interaction kinesics', in *HRI '08: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pp. 17–24, New York, NY, USA, (2008). ACM.

[47] S. Sabanovic, M. Michalowski, and R. Simmons, 'Robots in the wild: observing human-robot social interaction outside the lab', in *Proceedings of the 9th International Workshop on Advanced Motion Control (AMC 2006)*, pp. 596 – 601. IEEE, (March 2006).

[48] B. Scassellati, 'Building behaviors developmentally: A new formalism', in *1998 AAAI Spring Symposium "Integrating Robotics Research"*. AAAI, (1998).

[49] B. Scassellati, 'Theory of mind for a humanoid robot', *Auton. Robots*, **12**(1), 13–24, (2002).

[50] B. N. Schenkman and F. U. Jonsson, 'Aesthetics and preferences of web pages', *Behaviour & Information Technology*, **19**(5), 367–377, (2000).

[51] K. R. Scherer, 'Vocal communication of emotion: A review of research paradigms', *Speech Communication*, **40**, 227–256, (2003).

[52] J. Scholtz, M. Theofanos, and B. Antonishek, 'Theory and evaluation of human robot interactions', in *36th International Conference on Systems Sciences*. IEEE, (2003).

[53] H. Sharp, Y. Rogers, and J. Preece, *Interaction Design: Beyond Human Computer Interaction*, Wiley, March 2007.

[54] T. Shibata, M. Yoshida, and J. Yamato, 'Artificial emotional creature for human-machine interaction', *Systems, Man, and Cybernetics, 1997. 'Computational Cybernetics and Simulation'., 1997 IEEE International Conference on*, **3**, 2269–2274 vol.3, (Oct 1997).

[55] T. S. Shikler, *Le ton fait la musique - Analysis of expressions in speech*, Ph.D. dissertation, University of Cambridge, 2007.

[56] T. S. Shikler and P. Robinson, 'Recognising emotion in speech for human-computer interaction', in *International Workshop on Universal Access and Assistive Technology*, pp. 153–162, (March 2004).

[57] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh, and C. Rich, 'Explorations in engagement for humans and robots', *Artif. Intell.*, **166**(1-2), (2005).

[58] K. Stubbs, P. J. Hinds, and D. Wettergreen, 'Autonomy and common ground in human-robot interaction: A field study', *IEEE Intelligent Systems*, **22**(2), 42–50, (2007).

[59] M. Thuring and S. Mahlke, 'Usability, aesthetics and emotions in humantechnology interaction', *International Journal of Psychology*, **42**(4), 253–264, (August 2007).

[60] N. Tractinsky, A. S. Katz, and D. Ikar, 'What is beautiful is usable', *Interacting with Computers*, **13**, 127–145, (2000).

[61] G. Trafton, N. Cassimatis, M. Bugajska, D. Brock, F. Mintz, and A. Schultz, 'Enabling Effective Human-Robot Interaction Using Perspective-Taking in Robots', *IEEE Transactions on Systems, Man, and Cybernetics*, **35**(4), (2005).

[62] D. Vernon, G. Metta, and G. Sandini, 'The icub cognitive architecture: Interactive development in a humanoid robot', in *6th International Conference on Development and Learning (ICDL)*. IEEE, (July 2007).

[63] K. Wada and T. Shibata, 'Living with seal robotsits sociopsychological and physiological influences on the elderly at a care house', *IEEE Transactions on Robotics*, **23**(5), 972–980, (Oct. 2007).

[64] W. Wallach and C. Allen, *Moral Machines: Teaching Robots Right from Wrong*, Oxford University Press, 2008.

[65] R. D. Ward and P. H. Marsden, 'Affective computing: problems, reactions and intentions', *Interacting with Computers*, **16**(4), 707–713, (August 2004).

[66] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, 'A survey of affect recognition methods: Audio, visual, and spontaneous expressions', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **31**(1), 39–58, (January 2009).

[67] P. Zhang and N. Li, 'The importance of affective quality', *Communications of the ACM*, **48**(9), 105–108, (2005).

The Negative Attitudes towards Robots Scale and Reactions to Robot Behaviour in a Live Human-Robot Interaction Study

Dag Sverre Syrdal, Kerstin Dautenhahn, Kheng Lee Koay, Michael L. Walters,[1]

ABSTRACT

This paper describes the use of the Negative Attitudes Towards Robots Scale (NARS) to explain participants' evaluations of robot behaviour styles in a Human-Robot Interaction (HRI) study. Twenty-eight participants interacted with a robot in two experimental conditions in which the robot's behaviour was varied. Reliability analysis and a PCA was performed on the NARS items, creating three new subscales. Correlations between the subscales and other evaluations of the robot's behaviour found meaningful results, supporting the use of the NARS in English speaking samples.

1 INTRODUCTION

The LIREC (LIving with Robots and intEractive Companions) project aims to investigate theoretical aspects of artificial companions and test these across a wide range of embodiments in different contexts and environments [1].

While research in the LIREC project has a strong technological component, the development of tools for standardised measurements of the quality of interactions, perceptions of robots and agents and pre-existing attitudes towards such agents, is very important in order to establish a means of objectively comparing the success of agent and robot behaviour across a wide range of interactions and embodiments.

Dautenhahn [2] suggests that current work in human-robot interaction(HRI) is characterised by heterogeneity, both in terms of methodologies and measurements used to study technologies and their impact. This allows the field of HRI to accumulate a wide range of information regarding the use and perceptions of different robots in different contexts, but also limits the replicability of results across the field as a whole. It is understood that the technology driven nature of the field, and as such, the need to evaluate specific technologies in specific contexts, is a motivating factor for conducting such research. It is, however, important to note that this may become a problem for the HRI community in the long-term as the lack of common benchmarks and measures may hamper communication and application of results across different research groups and projects, and thus the advancement of the field as a whole.

The multidisciplinary nature of the LIREC consortium ,as well as the qualitative differences between agent embodiments and use-contexts across the project, suggests that this issue is not only of particular interest, but also provides a unique opportunity to study the efficacy of measures intended to measure attitudes to robots and their behaviours across diverse situations.

2 THE NEGATIVE ATTITUDES TOWARDS ROBOTS SCALE

In order to establish an understanding of how the behaviour and embodiment factors of a robot are perceived and responded to by potential users, such tools will also be needed to establish an understanding of the idiosyncratic factors in the individual user that may impact on behaviour, as well as subsequent evaluations of a given interaction. We have previously [2-5] suggested that individual differences in terms of underlying personality, gender and demographics may play an important role. However, these effects may not necessarily translate into analogous behaviour across different interactions. Also, measures regarding experience with computers and robots have been used [3, 6]. The results from Walters et al. [6] in particular, suggests that the relationship between individual differences and evaluation of robots and their behaviours can be quite complex. While these measures may provide possible answers to explain participant responses to robots, sometimes they are heavily influenced by the context of use, as suggested by Mutlu & Forlizzi [7]. They found that while overall computer use was not relevant, using computers for playing games did have an impact. As such, pre-existing biases and attitudes towards robots are difficult to extrapolate purely from demographics, personality and a history of technology usage.

One such scale is the Negative Attitudes towards Robots Scale (NARS). The NARS was developed using a lexical method, in which its developers created a scale based on free-form responses from participants regarding anxieties towards robots (Nomura and Kanda 2003). This later formed the NARS

[1]Adaptive Systems Research Group, School of Computer Science, AL10 9AB Email: (d.s.syrdal, k.dautenhahn, k.l.koay, m.l. walters, }@herts.ac.uk.

Table 1 NARS Items with Subscales

Item No.	Questionnaire Item	Sub-Scale
1	I would feel uneasy if robots really had emotions.	S2
2	Something bad might happen if robots developed into living beings.	S2
3	I would feel relaxed talking with robots*	S3
4	I would feel uneasy if I was given a job where I had to use robots.	S1
5	If robots had emotions I would be able to make friends with them.*	S3
6	I feel comforted being with robots that have emotions.*	S3
7	The word “robot” means nothing to me.	S1
8	I would feel nervous operating a robot in front of other people.	S1
9	I would hate the idea that robots or artificial intelligences were making judgements about things.	S1
10	I would feel very nervous just standing in front of a robot.	S1
11	I feel that if I depend on robots too much, something bad might happen.	S2
12	I would feel paranoid talking with a robot.	S1
13	I am concerned that robots would be a bad influence on children.	S2
14	I feel that int the future society will be dominated by robots.	S2

(*inverse item)

Scale, and was used successfully to explain differences in participants' behaviour in live Human-Robot Interaction (HRI) studies [8, 9]. Nomura et al. [10] also examined the relationship between the NARS scale and the Robot Anxiety Scale (RAS) with participants' behaviour in a HRI trial. The NARS has also been proposed as a means of gauging changes in attitudes towards robots over time as a result of prolonged interactions [11]. These possible applications make the use of such a scale interesting for both general HRI research across different robot types, as well as studying how prolonged relationships with robot companions may influence potential users' attitudes towards robots.

The items in the NARS are presented in Table 1, along with the sub-scales they are assigned to. The sub-scales are as follows:

Sub-scale 1: Negative Attitudes toward Situations and Interactions with Robots
Sub-scale 2: Negative Attitudes toward Social Influence of Robots,
Sub-scale 3: Negative Attitudes toward Emotions in Interaction with Robots

An English translation of the items in the scale has been created using appropriate methods of translation and backwards translation to achieve a linguistically valid scale. However, this translation has primarily been used for the purpose of evaluating cultural differences in attitudes towards robots [12]. Findings from these studies have been counter-intuitive, and suggest that Western participants are more well-disposed towards robots than Japanese. Studies using other means of measuring such as Implicit Association Tests and non-standardised Likert-scale questionnaires have provided conflicting results suggesting that Western participants find robots more threatening than their Japanese counterparts [13].

While this issue could be seen as a threat to the overall validity of using the NARS with a non-Japanese population, researchers should also consider some of the inherent dangers of using standardised questionnaires for such cross-cultural evaluations. In the field of individual differences, there exists a body of research that suggests that comparing culturally different samples using only participants' scores on such a scale is problematic. It appears more appropriate to investigate how differences in behaviour or related attitudes within the samples can be explained by such scores [13]. Secondly, while the NARS translation into English is valid from a linguistic perspective, cross-cultural differences not related to language may alter the internal reliability of both the scale as a whole as well as its sub-scales [15]. As such, reliability analysis of both the scale and its sub-scales would be useful when applied to a non-Japanese sample.

This paper describes the use of the NARS in a live HRI study, in order to examine the internal validity of the scale and its sub-scales, and to ascertain if it can account for differences in reactions to a robot's behaviour amongst participants.

Table 2 Robot Behaviour Styles:

Behaviour	Robot A	Robot B
Path	Straight	Circuitous Route with respect to participant’s pose
Speed	Fast	Slow when close to participant
Camera	Static and Forward-facing	Moving and Tracking
Negotiation of Space	“Excuse me”, and continuing as soon as possible	“After you”, continues after participant has moved away
Initiative in Bringing Pen	Does not wait for participant	Waits for participant to look for /ask for pen
Initiative in Delivering Pen	Bringing basket with pen to side of table, close to participant putting it down	Waiting in front of table facing participant, waiting for the participant’s notice, then putting basket down.

3 METHOD

Twenty eight (14 male, 14 female; aged between 18-55) participants were recruited for the study from students and staff at the University of Hertfordshire from a variety of disciplines. These participants took part in two interaction sessions with a robot. The sessions took part in a seminar room that was transformed into a simulated 'living room' for the purpose of this study (Fig. 1). In both interaction sessions, the participants were asked to perform a task which involved moving in a shared space with the robot as well as requiring a pen which was brought to the seated participant by the robot.

The robot's behaviour would differ between the two sessions. These behaviours were labelled Socially Ignorant (A) and Socially Interactive (B). The main differences between these behaviours can be found in table 2, both in terms of shared spaces as well as other interactional differences. The two behaviour styles were defined by the research team in terms of how much the robot adjusted its behaviour to the participant, rather than treating her as any other obstacle in the environment.

Participants were invited to evaluate the robot's behaviour after each interaction session, as well as rate the robot on a personality scale.

The whole experimental session, including the questionnaires took about 45 minutes per participant.

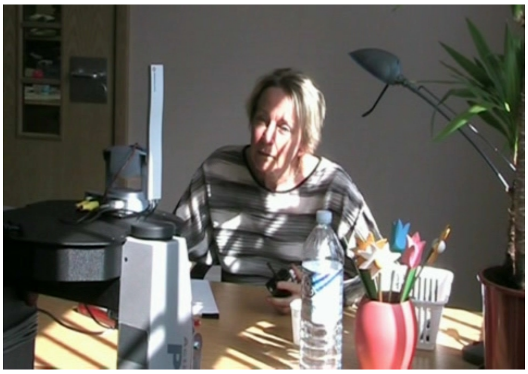


Fig. 1: A participant interacting with a robot, a Peoplebot robot (ActivMedia Robotics).

4 RESULTS

Results from these trials not related to the NARS scale can be found in [16], which discusses the relationship between participant personality traits and those attributed to the robot.

NARS Analysis

After administering the NARS scale to the participants, a reliability analysis using Cronbach's Alpha was performed on the participant's responses, and three items were removed from the analysis. These items were:

- The word 'robot' means nothing to me (S2)
- I would feel nervous operating a robot in front of others (S1)
- I feel that in the future, society will be dominated by robots (S2)

After these items were removed, the revised NARS scale had a Cronbach's Alpha of .80, supporting the notion of the scale as measuring a uni-dimensional construct.

However, [15] suggests that there were some cultural differences in how the scale functioned for a Western European sample. This would make an exploratory Factor Analysis using the Principal Components Analysis (PCA) method appropriate for investigating how attitudes towards robots would load within this sample.

The results from the PCA using the Varimax rotation method are in Table 3.

- **Future/Social influence -** had clear similarities with the Sub-scale 2 reported by Nomura et al.[9]
- **Relational attitudes** - which included items from all of the original Japanese sub-scales.
- **Actual interactions and situations -** shared characteristics with Sub-scale 1 in the original Japanese version (as opposed to the largely hypothetical aspects of the two other sub-scales highlighted by the double loading on one of the items).

Table 3: Subscale loadings

Item	Subscale		
	1	2	3
I feel that if I depend on robots too much, something bad might happen	0.86		
I am concerned that robots would be a bad influence on children	0.65		
I would hate the idea that robots or artificial intelligences were making judgements about things	0.54		
I would feel uneasy if robots really had emotions		0.81	
I feel comforted being with robots that have emotion*		0.79	
I would feel relaxed talking with robots*		0.75	
If robots had emotions I would be able to make friends with them*		0.64	
I would feel paranoid talking with a robot		0.44	
I would feel very nervous just standing in front of a robot			0.79
I would feel uneasy if I was given a job where I had to use robots			0.67
Something bad might happen if robots developed into living beings	0.48		-0.51

*inverse item

While there were similarities between these sub-scales and the original sub-scales reported by Nomura et al. [9], the differences between the two versions were considered large enough to merit the use of these sub-scales in this study.

This use should be considered tentative and was primarily done in order to understand the relationship between responses on the NARS and responses to robot behaviour styles within this sample of participants. It should be considered a response to the results from the Reliability Analysis which suggested differences in how the scale performed when compared to the results from Nomura et al. [9].

The revised NARS, as well as it sub-scales, was then used in a series of correlations in order to examine how they impacted evaluations of the robot's behaviour in each condition across the different types of interactions that took place in the experiment.

Participant Evaluation of Robot Behaviour Style
The evaluations the participants were asked to make were reported ratings on 5-point Likert scales when:

- Comfort for approaching or being approached by the robot.
- Comfort when physically close to the robot.
- Comfort moving in the same room as the robot.
- Comfort interacting with the robot whilst seated at the table.
- Reported overall enjoyment of the interaction.

A series of paired t-tests found that overall there were no significant differences between the two robot behaviour styles in how they were viewed by the sample.

Also, participants were invited to rate the robot according to 12 different traits suggested by Eysenck [17], anxiety, tension, shyness, emotional vulnerability, sociability, general activity level, assertiveness, excitement-seeking, dominance), and aggressiveness, impulsiveness, creativity. Note that some traits were removed from those originally suggested as they were considered unsuitable for describing robot behaviour. In addition, autonomy, controllability, predictability and considerateness were additional traits added. This was based on research that suggested that mental models of robots, while incorporating aspects of how we view humans, also include aspects that are defined by the robots' mechanical nature [18].

The Impact of the NARS
The focus of the subsequent analysis was to investigate if the NARS could be used to differentiate participant evaluation of the robot, both in terms of how participants evaluated the robot's behaviours and how participants attributed personality to the robot.

Evaluating Robot Behaviour Styles
In order to investigate the relationship between NARS scores and post-experimental evaluations of robot behaviour, a series of correlations were run between the NARS and its sub-scales with participant evaluations of the robot's behaviour styles.

The most important trait for evaluating robot behaviour styles within the interaction was the third tentative sub-scale: **Actual Interactions**. In terms of Robot A (Socially Ignorant) it impacted on both Comfort when being physically close to the robot (r*(28)=.448, p=.02*) as well as comfort when approaching or being approached by the robot (*r(28)=.464, p=.01*). For Robot B (Socially Interactive), the **Actual Interactions** sub-scale also impacted Comfort when being physically close to the robot (r*(28)=.442, p=.02*) as well as comfort when approaching or being approached by the robot (*r(28)=.466, p=.01*). It also impacted Comfort when moving in the same room as the robot (*r(28)=.462, p=.01*) and the Overall enjoyment of the interaction (*r(28)=.393, p=.04*)

The Overall NARS scores had no significant relationships with evaluations of Robot A's behaviour. There were however significant relationships between the NARS and Comfort when interacting with the robot while seated at the table (*r=(28)=.425, p=..02*) and Overall enjoyment of the interaction (*r(28)=.383,p=.04)*.
Note that for all these significant correlations a higher score on the NARS or its sub-scale, suggests a less favourable evaluation of the interaction.

Table 4 Correlation between trait attributions and NARS scores for Robot A (Socially Ignorant)

Trait	Overall Nars	Social/Future Implications	Emotional Attitudes	Actual Interac.
Anxiety	r=-.476, p=.010		r=-.436, p=.02	
Tension				
Shyness	r=-.441, p=.02			
Emotional Vulnerability				
Sociability				
General Activity Level				
Assertiveness				
Excitement Seeking	r=-.430, p=.022		r=-.394, p=.04	
Dominance				
Aggressiveness				
Impulsiveness				
Creativity			r=-.377, p=.05	
Autonomy		r=.389, p=.04		
Controllability				
Predictability				
Considerateness				

These results suggest that the NARS differentiated between participant responses to the two different robot behaviour styles. Participants with a higher score on the NARS scale found Robot B's behaviour less comfortable across a wider range of interaction sequences.

Attributing Traits to the Robot

In order to investigate the relationship between participants' NARS scores and how traits were attributed to the robot, a series of tests were run, correlating the NARS and its sub-scales with the different traits. The results from these correlations are shown in tables 4 and 5.

Table 5 Correlations between trait attributions and NARS scores for Robot B (Socially Interactive)

Trait	Overall NARS	Social/Future Implications	Emotional Attitudes	Actual Interac.
Anxiety		r= -.491, p=.01		
Tension				
Shyness		r= -.434, p=.02		
Emotional Vulnerability				
Sociability				
General Activity Level				
Assertiveness				
Excitement Seeking	r= -.446, p=.02			
Dominance				
Aggressiveness				
Impulsiveness				
Creativity				
Autonomy				r=.407, p=.03
Controllability	r= -.402, p=.03		r= -395, p=.04	
Predictability				r= -.384, p=.04
Considerateness				r= -.457, p=.01

Summarising these results presented in table 4 and 5, the NARS and it sub-scales correlate significantly with the traits attributed to both Robot A and Robot B. However, the overall picture emerging from these correlations is less clear than that for the evaluation of the interactions.

In general, it seems that Negative Attitudes Towards Robots Scale and its sub-scales are associated for both robots in terms of seeing the robot as the less anxious, less shy and less excitement seeking. What is more interesting for this particular investigation are the correlations between NARS scores and the robot specific traits. For these traits, the NARS sub-scale of **Actual Interactions** serves to differentiate between how participants attribute traits to the two different robots. Participants having higher scores on the **Actual Interaction** sub-scale tend to rate Robot B as more autonomous, less predictable and less considerate. Also higher **Overall NARS** and **Emotional Attitudes** were associated with seeing Robot B as less Controllable. This relationship is not seen for Robot A.

5 DISCUSSION

The results suggest that using the English translation of the NARS is an appropriate method of investigation prior attitudes towards robots that may impact participant evaluations of robot behaviour styles. After assessing the NARS using the Cronbach's Alpha as a measure of internal consistency, and removing three items, it had a high degree of internal consistency in a sample recruited at a British University. As suggested by Auer et al.[15], when using standardised measures across cultures, certain artefacts that originate in particulars of a given culture may impact both internal consistency as well as the validity of such a measure when applied to other cultures. It is possible that the three items removed may originate in such artefacts, specific to Japanese culture. This can also may serve as a possible explanation as to the differences in how the PCA performed in the responses from this sample described the sub-scales and those suggested by Nomura et al. [9]. We propose that such artefacts, rather than actual differences in cultural attitudes towards robots, may be the cause for the divergence of results from this study with those performed on Japanese samples. MacDorman et al. [13] suggests that these differences are not as pronounced as they are often believed to be.

More importantly however, is the utility of the NARS to explain, and possibly predict, other aspects of how people view and evaluate robot behaviour styles. In terms of differentiating between the two types of robot behaviours in this study, use of the NARS and its sub-scales differentiated between robot behaviour styles, which over the sample as a whole were not evaluated differently. Of particular interest is that higher scores on the NARS and the Actual Interactions sub-scale were associated with a more negative evaluation of the behaviour of Robot B, which was actually intended to act in a more socially appropriate manner.

There may be several reasons for this. One reason may be that this robot was seen as more socially sophisticated and that participants scoring high on the NARS as well as the sub-scale may be more wary of robots displaying a higher degree of sophistication.

Another explanation can be found in relating the trait attributions to these evaluations. It appears that participants with higher scores in the Actual Interactions sub-scale were more likely to rate Robot B as more autonomous, and less predictable. This may have been caused by the robot's behaviour. The behaviours by the socially interactive robot, could by some participants be considered more intrusive. Some of the behaviours of the robot, such as the movement of the camera, responding to the participants presence in terms of movement, waiting for participants to respond before leaving the pen, could have drawn attention to the robot as an autonomous agent within the scenario to a larger extent than Robot B's behaviour. This effect may be analogous as to that reported by Rickenberg & Reeves [19] when examining the impact of the behaviour of an animated character, in which participant evaluation of two differend behavioural styles varied dramatically depending on the participants' locus of control.

6 CONCLUSIONS

These results suggest that the Negative Attitudes towards Robots Scale may be susceptible to cultural differences. This may necessitate that research using this scale on a population outside of Japan may need to re-validate the scale and its sub-scales. However, our research also validates the value of the NARS as a means of explaining variance within a given sample in terms of evaluations of robot behaviour styles in a live HRI trial. Both the NARS and its sub-scales had an impact, not only on how participants evaluated their interactions with the robot, but also had some power to explain how participants differentiated between the two different robot behaviour styles. Negative Attitudes towards robots tended be associated with more negative evaluations of the behaviour of robot B (Socially Interactive behaviour style).

Of particular interest here, is that the sub-scale that had the strongest relationship with evaluations, **Actual Interactions**, might be considered to be related to the notion of robot anxiety as described in [11] as these items do refer to anxieties in actual interactions.

We would however, like to qualify the results from the PCA, as the number of participants was quite low due to resource constraints when running a live HRI experiments (the current study already took 2 months with daily HRI trials). However, this does not invalidate the meaningful relationships between at least one of the sub-scales and participant evaluations of the robots behaviour that were found.

7 ACKNOWLEDGEMENTS

We would like to thank our colleague Christina Kaouri for her help in setting up and running the experiments. The work described in this paper was conducted within the EU Integrated Projects, COGNIRON ("The Cognitive Robot Companion") and LIREC (LIving with Robots and integrated Companions). Funded by the European Commission under FP6-IST and FP7-ICT under contracts FP6-002020 and FP7-215554.

8 REFERENCES

[1] LIREC, http://www.lirec.org, 2008.

[2] K. Dautenhahn, "Methodologies and Themes of Human-Robot Interaction - A Growing Research Field," *International Journal of Advanced Robotics Systems*, vol. 4, pp. 103-108, 2007.

[3] D. S. Syrdal, K. Dautenhahn, S. Woods, M. Walters, and K. L. Koay, "Looking Good? Appearance Preferences and Robot Personality Inferences at Zero Acquaintance," *Multidisciplinary Collaboration for Socially Assistive Robotics: Papers from the AAAI Spring Symposium -Technical Report SS-07-07*, pp. 86-92, 2007.

[4] D. S. Syrdal, K. Dautenhahn, S. Woods, M. L. Walters, and K. L. Koay, "'Doing the right thing wrong' – Personality and tolerance to uncomfortable robot approaches," *Proc. 15th IEEE International Workshop on Robot and Human Interactive Communication(RO-MAN 2006)*, pp. 183-188, 2006.

[5] D. S. Syrdal, K.-L. Koay, M. L. Walters, and K. Dautenhahn, "A personalised robot companion? The role of individual differences on spatial preferences in HRI scenarios.," *IEEE International Symposium on Robot and Human Interactive Communication(Ro-man), Jeju Island, Korea*, 2007.

[6] M. Walters, D. S. Syrdal, K. Dautenhahn, R. T. Boekhorst, K. L. Koay, and S. Woods, "Avoiding the Uncanny Valley – Robot Appearance, Personality and Consistency of Behavior in an Attention-Seeking Home Scenario for a Robot Companion," *Autonomous Robots*, vol. 24, pp. 159-178, 2008.

[7] B. Mutlu, J. Forlizzi, and J. Hodgins, "A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior," presented at Humanoid Robots, 2006 6th IEEE-RAS International Conference on, 2006.

[8] T. Nomura, T. Kanda, T. Suzuki, and K. Kato, "Psychology in Human-Robot Communication: An Attempt through Investigation of Negative Attitudes and Anxiety toward Robots," *Proc. 2004 IEEE International Workshop on Robot and Human Interactive Interactive Communication, Kurashiki, Okayama Japan, September 20-22, 2004*, pp. 35-40, 2004.

[9] T. Nomura, T. Kanda, and Y. Suzuki, "Experimental Investigation into Influence of Negative Attitudes toward Robots on Human-Robot Interaction," *AI&Society*, vol. 20, 2006.

[10] T. Nomura, T. Suzuki, T. Kanda, and K. Kato, "Altered attitudes of people toward robots: Investigation through the Negative Attitudes toward Robots Scale," *Proc. AAAI-06 Workshop on Human Implications of Human-Robot Interaction*, pp. 29-35, 2006.

[11] T. Nomura, T. Kanda, T. Suzuki, and K. Kato, "Prediction of Human Behavior in Human--Robot Interaction Using Psychological Scales for Anxiety and Negative Attitudes Toward Robots," *Robotics, IEEE Transactions on*, vol. 24, pp. 442, 2008.

[12] C. Bartneck, T. Nomura, T. Kanda, T. Suzuki, and K. Kennsuke, "A cross-cultural study on attitudes towards robots.," *Proceedings of the HCI International 2005, Las Vegas.*, 2005.

[13] K. F. MacDorman, S. K. Vasudevan, and C.-C. Ho, "Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures," *AI & Society*, vol. 23, pp. 485-510, 2009.

[14] G. Matthews, I. J. Deary, and M. C. Whiteman, *Personality Traits*. Cambridge, UK: Cambridge University Press, 2003.

[15] S. Auer, H. Hampel, H.-J. Möller, and B. Reisberg, "Translations of Measurements and Scales: Opportunities and Diversities," *International Psychogeriatrics*, vol. 12, pp. 391-394, 2000.

[16] S. Woods, K. Dautenhahn, C. Kaouri, R. te Boekhorst, K. L. Koay, and M. Walters, "Are Robots Like People? - The Role of Subject and Robot Personality Traits in Human-Robot Interaction Trials," *Interaction Studies.*, vol. 8, pp. 281-305, 2007.

[17] Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach*. New York, USA: Plenum.

[18] S. Kiesler and J. Goetz, "Mental Models and Cooperation with Robotic Assistants," *Proceedings of the ACM CHI'02 Conference on Human Factors in Computing Systems*, pp. 576–577., 2002

[19] R. Rickenberg and B. Reeves, "The Effects of Animated Characters on Anxiety, Task Performance, and Evaluations of User Interfaces.," *Proceedings of the ACM CHI 2000 Human Factors in Computing Systems Conference.*, pp. 49-56., 2000.

The boy-robot should bark! – Children's Impressions of Agent Migration into Diverse Embodiments

Dag Sverre Syrdal, Kheng Lee Koay, Michael L. Walters and Kerstin Dautenhahn[1]

ABSTRACT

This paper presents results from a series of focused group discussions with a sample consisting of approximately 180 children during which views and opinions regarding agents migrating between different embodiments were elicited. The discussions attempted to ground the concept of a *migrating agent* in the children"s own experience of interacting with virtual characters in electronic toys and video games. The results suggest a complex interplay between expectations and appearance, and that disentangling the form of an agent may take from the underlying structures defining the agent's personality may be problematic for potential users.

1 INTRODUCTION

The aims of the LIREC (LIving with Robots and intEractive Companions) project [1] are to investigate the theoretical aspects of artificial companions and embody these aspects into robust and innovative technologies, both in the form of virtual agents as well as physical robots. This will allow for examining the applicability of these theories in actual social environments and facilitating the creation of artificial companions suitable for long-term interactions.

This endeavour includes studying both how a single agent can migrate into different embodiments depending on the tasks that it performs or the preferences of its users, as well as aspects of personalisation and adaptation to the particular idiosyncratic needs and preferences of diverse users.

While a major part of the project is to conceptualise, define and implement technological solutions to facilitate this process, it is also important to consider how prospective users may perceive and understand migration. Key questions for the LIREC project are how the unique underlying agent may be recognisable to the user in these different embodiments, as well as express personalised social behaviour when interacting with its users.

The importance of the use of affective and relational cues when creating and maintaining relationships between an agent and its users has been addressed by Bickmore et al. [2-4] who propose the use of such strategies and demonstrates the impact of their use with anthropomorphic virtual conversational agents.

Kasap et al. [5] propose an emotion engine which allows for emotive communication across different embodiments using a virtual anthropomorphic character and an anthropomorphic robot head. It also allows for episodic memory of previous interactions to be stored, facilitating long-term interactions and the formation of a long-term relationship between the agent and its interactants. While not addressing the topic of migration as such. they suggest that the affective communication ability in combination with the ability to retain memories of previous interactions are key in the development of relationships with artificial agents.

Martin et al. [6] propose an agent capable of migrating into diverse embodiments, and highlights the issue of agent perception in the user, and suggest that the changing nature of an agent's embodiment may degrade the visual cues that may be necessary for recognition and be an impediment to sustaining a continuous relationship. The use of persistent visual or behavioural cues is suggested as a means to counter this impediment. Their findings suggest that the use of visual cues is a powerful tool in aiding recognisability of the individual agent across diverse embodiments. However the use of other cues, as unique behavioural patterns or auditory cues are highlighted as issues that remain open.

2 THE PRESENT STUDY

This study explored children's perceptions of agent migration. Previous work in HRI has addressed how robots are perceived in terms of capabilities and moral agency by children [7, 8] and has indicated that children are capable of quite sophisticated reasoning regarding the nature of artificial entities. As such, insights from such a sample would be beneficial in the design and implementation of relational and affective behaviours as well as for identifying strategies for a migrating artificial agent. While the LIREC project does not focus on children in particular, products such as video games and electronic toys that target this demographic group often incorporate artificial, interactive agents. Therefore, this age-group is likely to have more everyday experience of interacting with such agents than an older population sample.

The topics of interest that were addressed were as follows:

1. Would a sample consisting of 8-10 year old children understand the concept of migration?
2. How would the relationship between the agent and its embodiment be considered?
3. How would this relationship be considered in light of the possibility of migration?

3 PROCEDURE

This study was conducted as part of a larger event in which children from local primary schools visited the Adaptive Systems Research Group at the University of Hertfordshire in

[1]Adaptive Systems Research Group, School of Computer Science, AL10 9AB Email: (d.s.syrdal, k.dautenhahn, k.l.koay, m.l. walters, }@herts.ac.uk.

May 2008. It was conducted as a series of group discussions in which one of the researchers would lead a discussion through which the children's impressions and ideas related to agent migration were elicited. There were a total of around 180 children participating in these discussions, in groups ranging from 3 to 6 children in any given group. While no attempt was made to balance the sample according to gender, all the students were from mixed-gender schools and the composition of the groups reflected this.

The discussion was conducted in a similar manner to a school class, where information was presented in order to facilitate further discussion. While the discussion was divided into stages, the researchers endeavoured to make the topic as responsive to the input of the children as possible, The stages of this discussion are reported below, with the particular questions that the children were encouraged to discuss in italics

1. Introduction to the notion of artificial agents:
   a. Highlighting the difference between characters in interactive media vs. films and books.
   b. Use of characters from video games (see Figure 1) as well as tamagotchis and other interactive toys.
      i. *Are these characters unique, what makes them different from one another?*

2. Introducing the concept of migration:
   a. Transfer of saved computer games to other consoles.
   b. Using electronic pets on websites.
      i. *Is it the same agent, even if it has moved to a different computer?*

3. Introducing the notion of migration into physical embodiment:
   a. Pictures of different robot embodiments, an iCub, a Sony AIBO and a Pioneer robot (see Figures 2-3).
      i. *What robot body would you prefer?*
      ii. *What robot body is the most useful?*

4. Introducing the notion of migration from one physical embodiment to another:
   a. Pictures of iCub and Sony Aibo shown;
      i. *How would you know it was the same agent?*

The above points were deliberately addressed in a loose manner, attempting to let input from the children drive the discussions during each stage as well as the introduction to the next stage.

The picture representing the notion of a migrating agent personality was a character in EA Game's MySims for the Nintendo Wii [9], see Fig. 1. This image was chosen for several reasons. First of all, the Wii console is very popular amongst the demographic that the presentations were given to. As such, the probability of one or more of the children in each group having experience of this game was quite high. Secondly, the notion of a MySims character being recognisable as an individual entity, both in terms of appearance, choice of activities and having a unique interaction history, both with the user as well as other in-game characters, is easily conveyable. The researchers would then use the sharing of experiences by some of the children in each group as a launch point for a discussion around the notion of agent personality. The initial grounding of the concept of an artificial agent, within the sphere of everyday experience of the children, was also intended to allow the children a greater repertoire for reasoning around the ideas that were explored in the discussion sessions.

Figure 1 Screenshot from EA game MySims for the Nintendo Wii used to exemplify a virtual character.

The images representing the robots were chosen in order to explore the large design space available to personal robots [10]. They offer three qualitatively different possible embodiments with three sets of equally different affordances. Having an anthropomorphic and zoomorphic robot as well as a clearly mechanical robot, facilitated exploration of a wide range of scenarios. This was both in terms of activities the participants could envisage the robots performing, as well as the nature of the interactions that would take place using these embodiments. Also, the researchers did not give out any information about the capabilities of these robots, so that any discussion regarding the use of the robots emerged from the capabilities the children projected unto them.

Following the discussion, the researcher demonstrated a form of migration where an agent 'personality' migrated between a Pioneer and a Peoplebot (see Figure 5). For the purposes of this demonstration, the personality was described to the children as being the way it avoided obstacles. The robot embodiments also used voice utterances as the agent migrated from one embodiment to the other.

During the demonstrations, one of the researchers took notes of the discussions and also noted interesting reactions to the demonstrations. These notes formed the core of the data to be analysed, but also served to highlight themes and issues that could be addressed in subsequent discussions with later groups. The discussions were also videotaped, with the consent of the participating schools as well as the guardians of the chidlren. There was considerable background noise, which made transcription and analysis of the raw video difficult.

4 RESULTS

The results from the discussions are described below. The focus of this analysis was primarily to explore how the children understood the role of an agent in different embodiments as well as migration. As such, the analysis presented is primarily descriptive in nature.

Figure 2 Picture of iCub [11] and SONY AIBO [12] shown during presentation

Artificial Agents

The main themes that emerged from discussing the notion of artificial agents were that of relating this to the children's own experience of video-games and other electronic toys. An interesting point here was that most of the groups explicitly made clear divisions between agents in computer games which are directly controlled by the player, and as such are extensions of the player, and agents that displayed different degrees of autonomy.

This was particularly relevant to how the children discussed the uniqueness of a given instance of a video game character. Most groups initially approached this in terms of the appearance of a character. However, probes from the researchers regarding behaviour were often associated with references to personality.

'Sims like different things, some Sims like to clean while others like playing more'

References to Tamagotchis tended to be linked with the possibility of the death of the agent. This particular feature of these electronic toys was in most groups associated with discussions of the uniqueness of the character emerging from a shared interaction history.

'You can start a new game with a new one…it is not the same. You haven't done anything with the new one…it doesn't know you.'

Migration from one Computer to Another:

The notion of a character in a video game being transferred from one computer/games console to another was not problematic to the sample. All the groups could easily volunteer means of doing so, including email transfer of game data as well as physically moving storage media from one place to the other, before connecting them to the new media. There was a general consensus in all the groups that the character would remain unchanged throughout this process.

Figure 3 Picture of Pioneer shown during presentation [14].

Migration into Physical Embodiment

Discussion centred around the groups' preferences as to what robot body the agent should inhabit. The majority of groups (likely due to having a bias towards game-like characters introduced earlier) focused on the play possibilities of the different embodiments. This led to a preference for the iCub and the Sony AIBO embodiments.

'The dog-robot looks like it can play.'

'The human looking one, because he can play games with me.'

Preferences for the AIBO were often justified in terms of it being dog-like, and reflecting an underlying liking of dogs in general, as well as a clear understanding of the play-possibilities with dogs that could be transferred to interactions with the AIBO embodiment.

'I like the robot dog…no reason, but I really like dogs.'

'I like the robot dog, because I have a dog and I play with it all the time, and we have fun together.'

'I like the robot-dog, it could run after balls and it would be fun.'

Likewise, the iCub was credited with human-like capabilities in terms of speech, as well as intelligence.

'The boy-robot could keep me company…we could talk'

'I would like it [the iCub] *to help me with my homework.'*

On the other hand, groups in which the discussions were led towards other tasks started to have more detailed discussions regarding the possibilities and limitations of each embodiment when executing specific tasks. A common task that was discussed by a large portion of these groups was that of fetching and carrying drinks or snacks. These discussions highlighted apparent affordances based on the images of the robot presented, both in terms of possibilities and limitations:

Figure 4 Group Discussion

'The human one has arms so he can lift things, and walk on his legs to bring you a drink'

'The pioneer-robot could bring you things and drive around.'

'The human one would catch fire if it got water on the wires; maybe it shouldn't use the tap.'

'The one with wheels doesn't have any arms, so it can't pick anything up.

Interestingly, the AIBO embodiment was only considered suitable for particular tasks that the participants considered appropriate for dogs to do:
'It [the AIBO] could get the newspaper.'

'The robot-dog could guard my things.'

Also, some of the groups started considering the possibilities of collaboration between the robot embodiments to better perform tasks. The following quote regarding a fetch and carry task serves to illustrate this:

'The human one[the iCub] *can't walk very fast…maybe it could put the glass on the one with wheels* [the Pioneer] *so it could bring it to you?...I have never seen a fast walking robot'*

Interestingly, some children considered the difficulty for the agent in terms of orienting itself to a new body. This line of reasoning concluded that the humanoid iCub would be the most suitable for the agent, which was represented by the MySims screenshot, due to similarities of form:

'I think the human one, this one has two arms and two legs [points to MySim screenshot]*, and so does the human robot. It doesn't have to learn anything new, so it is easier for it.'*

Migration from one Physical Embodiment to Another

This particular issue raised questions from the children related to how the agent might represent itself across different embodiments. Drawing upon the discussions of the previous sections, the majority of the groups had already considered the notion of a *persistent, unique agent, with a particular interaction history with its users*. Two particular themes emerged as to how the agent could/should signal its identity to an onlooker.

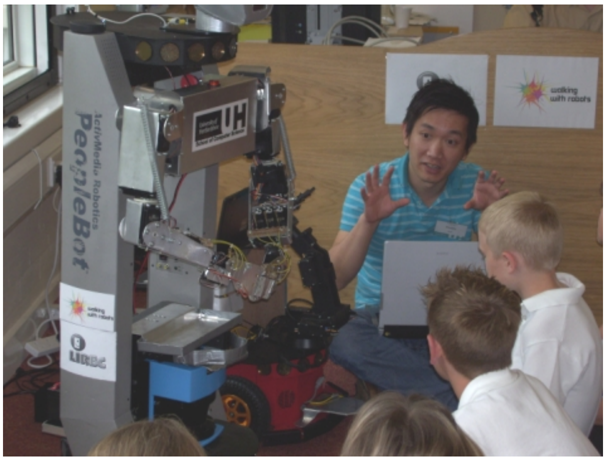


Figure 5 Migration Demonstration.

The first arose through reasoning which posited an original, ideal embodiment for the agent. This particular theme tended to incorporate an implicit assumption that the agent had a form which it spent the majority of its time in, and other embodiments were only adopted at a task-based basis. This led to suggestions of the robot adopting habits and behaviours that were clues to this original form in order to inform the user of its identity:

'If the character was in the dog and then moved to the boy-robot, then maybe the boy-robot should bark?…it would say woof woof!'

'Maybe the dog robot could walk on two legs?'

'When it is in the boy robot it would be very good at rolling over.'

'You would know that it has moved from the human one to the dog, because the dog robot could talk.'

It is interesting to note that participants did not consider such a transfer from the Pioneer embodiment. There was however, some comments that suggested such transfer from the iCub and the AIBO to the Pioneer.

The second theme that emerged followed a line of reasoning in which the group would see the interaction history and personality of the agent as something independent to the embodiments themselves. Following this argument, the groups would argue for analogous behaviours communicating affective behaviour.

'If the character is happy and in the dog it would bark and roll around…if it is in the boy, it could smile and laugh'

'If the character moves into the one with wheels it could spin around really fast if it is happy to see you.'

5 DISCUSSION

The findings from these focused group discussions suggest that children in this age-range are certainly capable of understanding the concept of agent migration into diverse physical embodiments. The use of examples and imagery from the children's everyday experience, through games and electronic toys, was particularly effective in eliciting meaningful responses from the participants.

Many of the responses from the children focused heavily on the play-aspect of such companions. This was to be expected due to initial focus on entertainment applications artificial agents in the slides used in the presentation.. Also, for this age-range most electronics products are intended as vehicles for entertainment. It is important to note that the participants did not have difficulty when prompted to consider applications other than play for the agents in different embodiments. Also, considerations such as engagement across different embodiments is still valid in interactions that are not intended as being solely for entertainment purposes [2].

This study was an exploratory study and the main focus was to gain a wide range of comments and insights into the relationship between how an agent is perceived in terms of its embodiment. Also, our aim was to examine how migration was perceived by the children, rather than examining specific pre-determined relationships between concepts. However, there were some interesting insights from the sample.

One of the most salient themes emerging in the discussions related to how the role of affordances, based on an embodiment, determined the role of the agent. This was in some instances based on the physical capabilities of the embodiment. For example, as in the discussions of whether to use iCub or the Pioneer for the fetch and carry task. However, the iCub and the AIBO embodiments also carried with them a set of expectations. These were not just related to apparent capabilities, but drew on expectations based on the form of the robot, wherein the robot would take on a social role based on what it appeared to be. Thus, fetching the newspaper, running after balls, and guard duty were considered appropriate tasks for the agent in the AIBO embodiment. Likewise, for the iCub, the ability to talk and help with tasks of a more intellectual nature was also considered appropriate

This was also reflected in the views of migration. In those discussions that posited an original form, the agent would retain the social and intellectual aspects of the role afforded to it by the original form. As such, identification of the unique agent would here be accomplished using cues that would hint at these roles, e.g. barking and rolling over if the migration was from the AIBO embodiment, or speaking if the migration was from the iCub.

A similar issue emerged in the statements of those groups who, when considering the best embodiment for the robot to take, decided upon the humanoid form of the iCub. The agent could then apply its knowledge about its virtual embodiment directly to that of the iCub.

These results can be considered in the light of previous work such as Walters et al. [15], which suggests that the behaviour of a robot should be consistent with the expectations created by its particular appearance. However, these results also suggest that adding migration to the mix might create a more complex and dynamic interplay between embodiment and expectations. The discussions suggested that behaviours could clarify an original set of affordances for the agent, *despite* those of its current embodiment.

It should be noted however, that some of the groups focused on the role of the agent as an entity divorced from its embodiment. These groups considered the various embodiments as avenues for interaction which the agent could use to express itself and act upon the world. However, these groups were in the minority and as such, the data on this reasoning is sparser.

6 CONCLUSIONS

This was an exploratory study and these results were not intended to be directly applicable to the implementation of migration processes of agents within the LIREC project. They are however, a source of future avenues of investigation.

The most prominent of these is the issue of how the agent initially should present itself. The power of a perceived 'ideal' embodiment for the agent should not be underestimated, both in terms of framing expectations as to (perceived) intellectual capabilities as well as its social role, As such, when initially presenting itself to the user, the form the agent is introduced in, might impact subsequent perceptions of the agent across different embodiments. This may be a powerful tool in terms of situating the role of the agent within the everyday experience of the user, especially if the social role afforded it by its embodiment is congruent with its capabilities. For instance, a robot intended for fetch and carry as suggested by [16] may benefit from being initially presented as having an original dog-like embodiment. Dogs are trained to perform such tasks for users and thus these affordances would then support the interactions resulting from these tasks. On the other hand, this may prove an obstacle to interactions if the agent is embodied in a form that can use different modalities to communicate than those which the user perceives in its original form. In which case, the user may find these modalities inconsistent with their expectations from the agent.

Therefore, examining the processes of how the perception of the agent's original or ideal embodiment is created by the user, as well as possible ways of shaping the creation of such a perception, may be useful in future work. Also, ways of utilising these perceptions are also an interesting avenue of investigation.

ACKNOWLEDGEMENTS

We would like to thank our colleague Cathy Beer for her help in compiling notes for analysis. The work described in this paper was conducted within the EU Integrated Project LIREC (LIving with Robots and intEgrated Companions). Funded by the European Commission under FP7-ICT under contract FP7-215554.

REFERENCES

[1] LIREC, http://www.lirec.org, 2008.

[2] T. W. Bickmore, L. Caruso, K. Clough-Gorr, and T. Heeren, "'It's just like you talk to a friend' relational agents for older adults," *Interacting With Computers*, vol. 17, pp. 711-735, 2005.

[3] T. W. Bickmore and J. Cassell, "Relational Agents: A Model and Implementation of Building User Trust," *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 396-403, 2001.

[4] T. W. Bickmore and J. Cassell, "Social Dialogue with Embodied Conversational Agents," in *Natural, Intelligent and Effective Interaction with Multimodal Dialogue Systems*, J. v. Kuppevelt, L. Dybkjaer, and N. Bernsen, Eds. New York: Kluwer Academic, 2005, pp. 23-54.

[5] Z. Kasap, B. Moussa, P. Chaudhuri, D. Hanson, and N. Magnenat-Thalmann., "From virtual characters to robots - a novel paradigm for long-term emotional human-robot interaction.," *Submitted to Human-Robot Interaction Conference, 2009.*, Submitted.

[6] A. Martin, G. M.P.O'Hare, B. R. Duffy, B. Schön, and J. F. Bradley, "Maintaining the Identity of Dynamically Embodied Agents," *The Fifth International Working Conference on Intelligent Virtual Agents - IVA 2005 Kos, Greece, September 2005*, 2005.

[7] P. H. Kahn, N. G. Freier, T. Kanda, H. Ishiguro, J. H. Ruckert, R. L. Severson, and S. K. Kane, "Design patterns for sociality in human-robot interaction," in *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*. Amsterdam, The Netherlands: ACM, 2008.

[8] S. Turkle, W. Taggart, C. D. Kidd, and O. Dasté, "Relational artifacts with children and elders: the complexities of cybercompanionship," *Connection Science*, vol. 18, pp. 347-361, 2006.

[9] S. Zhu, course materials for CMS.600 / CMS.998 Videogame Theory and Analysis, Fall 2007. *MIT OpenCourseWare (http://ocw.mit.edu/),* Massachusetts Institute of Technology. Downloaded 09-02-2009.

[10] K. Dautenhahn, 'Design spaces and niche spaces of believable social robots', Proceedings of the 11th Annual International Workshop on Robot and Human Interactive Communication (RO-MAN 02), Berlin, Germany, 192-197, (2002).

[12] Sony, "Sony AIBO," *http://support.sony-europe.com/aibo/*, 2009.

[13] RobotCub, *http://www.robotcub.org*, 2009.

[14] Pioneer, "Mobile Robots inc.," *http://www.activrobots.com/ROBOTS/p2dx.html*, 2009.

[15] M. L. Walters, D. S. Syrdal, K. Dautenhahn, R. te Boekhorst, K. L. Koay, 'Avoiding the Uncanny Valley – Robot Appearance, Personality and Consistency of Behavior in an Attention-Seeking Home Scenario for a Robot Companion', *Journal of Autonomous Robots*, 24(2), 159-178, (2008).

[16] H. Huettenrauch and K. Severinson Eklundh, "Fetch-and-carry with CERO: Observations from a long-term user study with a service robot.," *Proceeding of the 11th IEEE International Workshop on Robot and Human Interactive Interactive Communication(RO-MAN 2001), Berlin, Germany Sept 25-27, 2002*, pp. 158-163, 2002.

The Advantage of Mobility: Mobile Tele-operation for Mobile Robots

Alberto Valero[1] and **Gabriele Randelli**[2] and **Chiara Saracini**[3] and **Fabiano Botta**[4] and **Massimo Mecella**[5]

Abstract. Intra-scenario operator mobility is claimed to be a strong advantage when acquiring situational awareness within a robot tele-operation. This factor should not be discounted when seeking to build more effective Human-Robot Interaction (HRI) systems. In this paper, on the basis of extensive experimentation comparing a desktop-based interface wrt. a PDA-based interface for remote control of mobile robots, we provide support (and also some confutation) of this claim. The experiments were performed in order to identify the most suitable operator interface for controlling a mobile robot depending on the task and mobility/visibility of the operator.

1 Introduction

Let's suppose a team of robots is deployed in a nuclear plant to execute scheduled operations of surveillance and security controls. A nuclear plant is characteristically divided into different security areas. As one gets closer to the reactor the radioactivity increases and so also the risk of contamination for humans. The advantage of using robots in such situations is to avoid undesiderable risks for the workers while performing the inspection of the plant: robots go where humans fear to tread.

This working scenario would become critical if a nuclear accident, such as the explosion and subsequent fire of the Chernobyl Plant in the Soviet Union in 1986, were to occur. It is challenging for a first response team in accidents such as this to assess the extent of the damages and the associated risks. Emergency personnel deployed in a disaster zone normally cannot provide enough information about the state of the situation to the Center of Control and Operations (CCO) in order to plan the emergency response. The first responders who have reached the disaster zone usually are prevented from going beyond certain limits, due to high temperatures, radioactivity, or simply because they do not know the extent of the risks, being unable to assess the situation in its entirety. A robot team is again one solution in dealing with such hazardous scenarios, permitting the avoidance of unnecessary risks. Robots can be deployed to help the first responders to make a proper situation assessment. At the first moment operators would not have any visibility of the robot nor the scenario, as they cannot enter the disastered area. When an initial situation assessment is made, and areas safe for humans identified, responders (carrying hand-held devices) can go into the affected zones, having partial visibility of the robots and scenario. Robots can even guide responders under low visibility conditions to desired target points through safe paths [11].

As seen in this scenario, operators or responders must remotely drive robots into areas which they might not be seeing and that could be partially destructed. Remote driving of a robot in such conditions is not a simple process, but a multicomponential one. A successful navigation in an information rich space requires human cognitive abilities such as orientation, wayfinding, visuospatial representation of environment, planning, etc. When a human operator drives a robot through a Graphical User Interface (GUI), he or she must have a proper Situational Awareness (SA). The SA provided by the interface has been considered in literature one of the measures for evaluating its usefulness [4][12][13][14].

We are working on the design and implementation of interfaces thought for this kind of missions. In disaster situations or scheduled operations the human team is composed of on-site operators, which can only wear hand-held devices, and remote operators, having access to wider computerized systems. Even if remote operators, using powerful workstations, can visualize and process a wider amount of data, responders carrying a PDA interface can boost the pervasiveness of robotic systems in mobile applications, where operators cannot be fixed to a particular place. Even if mobile devices are less powerful than desktop computers, they offer the operator the capacity to move, allowing him to partially view the actual scenario with the robot that he is controlling. The disadvantage related to the device limitations could be balanced by the advantage of mobility. Mobility could grant better situational awareness enhancing the control of the robot. First responders can control a robot team with a PDA interface while having a partial view of the environment, and thus obtain on-field information not retrievable by the robot sensors. This is the advantage of mobility, which will be studied in this paper.

Recently, a growing interest is emerging in how to develop human oriented robotic interfaces [7] [12] [19] [3]. Such interfaces do not require as extensive a knowledge of the robot system while they permit the operator to control and/or supervise a robot or team of robots. Certain guidelines for developing interfaces usable for humans have been reported in [1]. This paper presents the results of an experiment comparing the usefulness of a PDA with a desktop interface in order to determine the optimal way of distributing the control of a robot between a mobile operator using a hand-held device and a stationary operator using a desktop computer [17]. Particularly, the main purpose was to investigate which of the two interfaces is more effective both in navigation and/or in exploration tasks, depending on the different conditions of visibility, possibility of operator movement, and environmental spatial structures. Our research question is: *may the*

[1] Department of Computer and Systems Sciences, SAPIENZA - Università di Roma, Italy, email: valero@dis.uniroma1.it
[2] Department of Computer and Systems Sciences, SAPIENZA - Università di Roma, Italy, email: randelli@dis.uniroma1.it
[3] Lab. of Cognitive Science and Psychology, SAPIENZA - Università di Roma, Italy, email: chiara.saracini@uniroma1.it
[4] Lab. of Cognitive Science and Psychology, SAPIENZA - Università di Roma, Italy, email: fabianobottaster@gmail.com
[5] Department of Computer and Systems Sciences, SAPIENZA - Università di Roma, Italy, email: mecella@dis.uniroma1.it

mobility of the operator inside the operating scenario counterbalance the disadvantages of a PDA device wrt. a desktop device?, and, if yes: *how?*, *under which circumstances and/or tasks?*. The anticipated final result of our research is to identify the situations and tasks which can counterbalance the limitations of the devices required by mobile operators and the circumstances in which the desktop interface is preferred even if the operator using it is fixed remotely.

The work is organized into four main sections. We begin giving some theoretical foundations from *spatial cognition sciences* [Section 3] in order to justify the preliminary hypothesis of the experiments [Subsection 4.3]. We continue showing our two interface prototypes: a PDA-based interface and a desktop based interface (which were used in the experiments) [Section 3]. Then, we describe the experiments and how the data were analyzed [Section 4]. Finally, we present the results and discuss the contribution of this study [Section 5] . A final section closes the paper by outlining future work.

2 Situational Awareness and Spatial Cognition

The commonly accepted SA definition was given by Endsley [6] and adapted to HRI by Yanco, Drury and Scholtz as *"the understanding that the human has of the location, activities, status, and surroundings of the robot; and the knowledge that the robot has of the human's commands necessary to direct its activities and the constraints under which it must operate"* [20]. This definition distinguishes three components within the concept of SA: human-robot SA, robot-human SA, and the human's overall mission awareness [20]. Within the human-robot awareness two aspects are important for the purposes of this paper: location awareness, defined as a map-based concept, allowing the user to locate the robot in the scenario, and surroundings awareness pertaining to obstacle avoidance, allowing the user to recognize the immediate surroundings of the robot [4].

In order to better understand how the SA enhance the operator performance when he or she is driving a robot is useful to introduce two important concepts from human spatial-cognition: *route knowledge* and *survey knowledge*. The distinction between route and survey knowledge helps to understand the cognitive skills required by a human operator remotely controlling a robot. The route perspective is closely linked to perceptual experience: it occurs under the egocentric perspective in a "retinomorphous reference system", that is, one is able to perceive himself in the space [9], with a special emphasis on spatial relations between objects composing the scene an agent is situated in. This is for example the case of an operator driving a robot with a tridimensional perspective on a screen, simulating the visual information that he or she would obtain by directly navigating in that environment (see Section 3.1, desktop interface 3-D viewer). Route-based information, from a ground perspective, is stored in memory to keep trace of turning points, distances and landmarks or relevant points of reference in the observed context. In contrast, survey perspective is characterized by an external and allocentric perspective, such as an aerial or map-like view, allowing direct access to the global spatial layout [2] as it would be if the operator had a device by which he or she can have a global, aerial view of environment and the robot inside it (see Section 3.2, PDA interface). Previous studies have shown that a navigator having access to both perspectives exhibit more accurate performances [9].

We can appreciate a relation between location awareness and survey knowledge, while surroundings awareness relates to route knowledge. Our case study consists of a human operator driving remotely a robot using a human-robot interface. When the operator is not physically in the navigation scenario, the interface must enhance his or her spatial cognitive abilities by offering multilevel information about the environment (route and survey). Complex interfaces can provide different perspectives of the environment (bird's eye view or first-person view). Such information allows an operator looking at a GUI (Graphical User Interface) to have more than one perspective at the same time. Contrarily, if the operator is in the scenario, part of the information can be acquired by direct observation, depending on the visibility the operator has. In such situations less information is required in the GUI.

These spatial-cognitive aspects should be taken in consideration when designing a human-robot interface for remote tele-operation. Unfortunately, HRI development tends to be an afterthought when designing robotic systems [1] and advances in AI, sensory fusion, path planning, autonomous navigation, image processing, etc. are not often integrated into proper interaction systems. Actually, most robot operation interfaces are system-oriented, permitting developers to have low level control of the system and facilitating the debug process but which are very difficult to operate for non-expert users.

3 Interface Prototypes

We have implemented two interface prototypes: one for desktop computers and one for PDA devices[6]. Both are based on the HRI interfaces discussed and analyzed by Yanco [4] and Nielsen [12]. Nevertheless, they only considered an egocentric point of view, either for video acquisition, either for map information. We associated to this approach an allocentric point of view, to enhance operator's SA, as discussed in Section .

3.1 The Desktop Interface

Our desktop interface is designed for controlling robots in structured and partially unstructured environments. Its scope is to be able to control a robot dealing mainly with exploration, navigation and mapping issues. Its main purpose is to enhance the operator's performance of complex tasks, with a comprehensive overview of the whole explored area, and supplying all the necessary tools to control the robot. The overall information is always visible on the screen while controlling the robot. The interface inplements also the possibility to control a team of robots.

The interface shown in Figure 1 can be divided into two parts: the topmost panel is the *Active Robots Panel*, where the user can switch among the robots of the team, in order to directly interact with an individual unit. If a robot is added to the team, the operator can easily connect with it. The rest of the window contains all the information relative to the selected robot and the robot team:

Navigation Panel. It is localized in the central area of the window, it is composed of a *Local View of the Map* and a *Global View of the Map* giving a bird's eye view of the zone. The map is constructed on-line from the robot laser range sensor and the odometry using SLAM (Simultaneous Localization and Mapping) techniques. The Local View can be zoomed in and out. The robot is located within the map by a rectangle-symbol containing a solid triangle that indicates its direction. The second component is the *3D Viewer*, which allows a more comfortable and realistic navigation in the tele-operation mode giving an egocentric perspective of the scenario. The pseudo-3D reconstruction may be based either on the laser range data or on the 2D map, by simply elevating the obstacles into 3D images. The laser view is more precise than the map view, but it only gives information of the obstacles in front of the robot, while the map view gives a

[6] They're both available at http://www.dis.uniroma1.it/~valero/sw/.

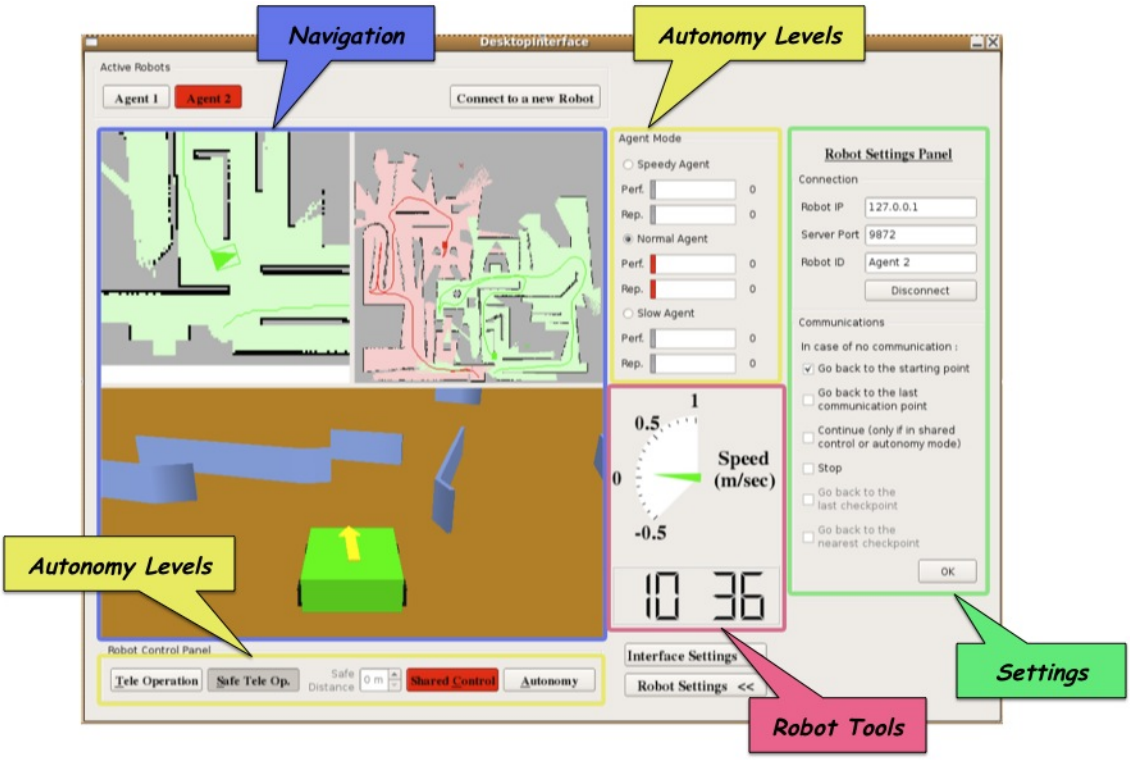


Figure 1. Desktop Interface

more global picture. The laser view is demonstrated to be more useful to drive the robot in narrow spaces in which the constructed map would not be adequately precise. The operator can manually switch from one to another as well as pan-tilt them.

Autonomy Levels Panel. It allows the operator to switch among four control modes: *tele-operation*, *safe tele-operation*, *shared control* and *autonomy*. In the safe tele-operation mode the system prevents the robot from colliding with obstacles. In the shared control mode the operator sets a target point for the robot by directly clicking on the map, which the robot tries to reach. When working in shared control or autonomy, the operator can select from three agents (*Agent Mode Panel*): *slow*, *normal* and *speedy*, which have different pre-set maximum velocities and use different heuristics to explore the environment.

Robot Tools Panel. This panel consists of several widgets to monitor the robot kinematics. There is a speedometer, a chronometer to keep track of the mission length, and a gyroscope directly embedded in the robot within the 3D View (yellow directional arrow).

Settings Panel. It is divided into two views and it consists of the *Interface Settings* and the *Robot Settings*.

3.2 The PDA Interface

Due to the reduced size of a PDA and its computational limitations, the display cannot present on-screen all the data provided by the HRI system. In order to keep the same functions offered by the desktop-based interface we implemented them using various simplified layouts. This underlines how critical it is to present the operator only the crucial data, as each layout change implies a longer interaction time with the device. Another critical point was to consider the slower input capacities of the operator with a PDA, consisting of a touch screen and a four-way navigation joystick. Thus, it is important to minimize the number of interactive steps to change a setting or to command the robot.

The PDA has two kinds of 2D views, each selectable with its own tab. The first, egocentric, is the *Laser View* (Figure 2(a)). The second is the *Map View* (Figure 2(b)), equivalent to the *Global Map View* described in 3.1. A third tab (Figure 2(c)) is dedicated to the *Robot Control* functionalities, merging both the *Autonomy Levels Panel* and the *Agent Mode Panel* of the desktop version. The interface and robot settings can be modified by clicking the tab located at the bottom of the display.

4 Experimenting with the interfaces

Figure 3. The P2AT robot inside the indoor area during one of the experiment runs

(a) *Laser View*. It offers a precise real-time local representation of what obstacles we are facing.

(b) *Map View*. It allows the operator to retrieve the map of the explored area and set a target point by clicking on the map.

(c) The *Autonomy Level Settings Window*, It allows the user to set his desired robot control mode.

Figure 2. PDA Interface.

There have been studies done on how different designs of an interface and the interaction model can support the operator when operating and supervising a remote robot(s). These studies compare the usefulness of various desktop-based interfaces [5][4][12] or of various PDA-based interfaces [10][7]; however, very few studies compare desktop-based wrt. PDA-based interfaces for remote robot control. Our interest was to determine which are the main performance differences between a PDA and a desktop-based interface.

4.1 Subjects

Twenty four subjects (four females and twenty males) ran the experiments, aged between 20 and 30, nineteen undergraduate and five PhD. candidates. The scenarios of the experiments were different and no participant had previous experience with any of the two interface prototypes. All of the participants completed the three experiments in in the same order and so on no one had more experience than the others. We looked for a trade-off between people with experience in robotics and computer science and people with no experience.

4.2 Experiment Design and Procedure

Three experiments were run. The whole experimental context was scheduled in five days. The subjects were splitted into two groups, one using the PDA interface, the other the desktop version. Every subject was trained for twenty minutes to acquire a basic knowledge of the functionalities provided by the interfaces. After the training, they ran the experiments in order. Each subject had a single trial.

First experiment The first experiment was run using Player/Stage [8] robotics simulator. Subjects were asked to explore a virtual unknown environment of 20 m x 20 m (Figure 5) using a mobile robot equipped with laser range scanner. Users were given twenty minutes of time to explore the maximum area without colliding. Each candidate was assigned a type of interface randomly and had a single trial with it (twelve subjects drove the robot using the PDA and the other twelve the desktop interface). In order to "motivate" this task they were asked to look for "radioactive sources" distributed in the area. The sources were detected by a simulated sensor installed on the robot. During every run, the operator was supported and supervised by one assistant, who previously helped him or her in the training. Another person was in charge of supervising the correct-functioning of all the software. Returning to the scenario application given in the introduction, this experiment applies to the case in which responders must explore a disastered area in order to asses sthe extent of damages after a nuclear accident.

The independent variable was the *Interface Type* $\in$ {*PDA interface, desktop interface*}, while the rest of the factors remained unchanged. The dependent variable was the covered area by the robot in square meters. An area was considered covered if it had been mapped.

Figure 4. Operator driving the robot with the PDA interface. The robot appears from a hidden area prior entering the building

Second experiment Subjects were asked to navigate with a real Pioneer P2AT robot equipped with a SICK Laser Range Finder alongside a path composed by narrow spaces, cluttered areas and corridors (the path was about 15 meters long). Users did not need to find a way but just follow it from the beginning to the end. Subjects were not given a layout of the scenario and it was never visible to them. During every run, the operator was supported and supervised by three assistants: one who trained him or her for five minutes to use the real robots and recorded some data during the trial. The second person was the technical responsible of the robot and of the interface device, while the last controlled the robot during its motion and took care of its safety and of the scenario. This second experiment tries to reproduce the situation in which operators must drive remotely a robot to a target point, like in scheduled operations in nuclear plants.

The independent variable was the *Interface Type*. As for the dependent variable, we were interested in the *Navigation Time*, measured in seconds.

Third experiment. Subjects were asked to navigate again in a real scenario with the same P2AT robot. The environment consisted of an outdoor area in a courtyard, linked through a ramp to a corridor inside our department. The desired scenario recalls a disastered area and is composed by three different zones, all realized using reclining panels and cartons:

- *Maze*, with one entrance and one exit;
- *Narrow Spaces*, very tight areas where the robot can only pass through them, without any chosen direction;
- *Cluttered Areas*, contain several obstacles placed irregularly and isolated in the area, such that the robot can navigate through the area choosing more than one direction.

In this last experiment subjects using the PDA could move in the scenario, resulting in situations in which they could completely see the scenario and robot, and areas in which they could just see them partially. The outdoor area was visible to the operator while the robot was non always completely visible. Operator could not enter the *arena*. The indoor area was only partially visible through some windows located at the top of the indoor scenario and the robot was completely hidden at least half of the path. Outdoor and indoor areas were different and measured times cannot be compared among them. The operator was supported like in the previous experiment. This last experiments simulates a nuclear disaster after the initial situation assessment. Responders know where they can go without contamination risks, and thus, responders carrying a PDA can drive the robot partially seeing it. There will be some limits that responders will not be able to trespass, and thus, they will have to drive the robot with no visibility of it.

We used a 3x2x2 factor design, where the independent variables were: *Space Type* $\in$ {*Maze, Narrow Spaces, Cluttered Areas*} for the part of the path, *Operator View Degree* $\in$ {*Total Visibility, Partial Visibility*} which determines the operator direct view of the scenario and robot, and finally *Interface Type*. The first two were treated as "within-subjects" factors, while the last one as a "between-subject" factor. The measured variable was the time (in seconds) required to complete the path.

The scenario, the robot configuration, and the wireless signal strength were the same for all the subjects, to guarantee replicability.

4.3 Preliminary Hypothesis

The different technical features of the desktop and PDA interfaces are strictly connected with different ways of acquiring information from the navigated space and consequently with different behavioral capacities to build a mental map of the environment depending on the scenario features. The laser and the map views, which are available in both desktop and PDA interfaces, represent respectively "route" and "survey" knowledge [15][16], as defined in the introduction.

These perspectives convey respectively to path-planning and survey knowledge. Path-planning, in order to avoid obstacles, depends on the operator surroundings awareness; spatial information is accessed sequentially, the number of paths emanating from each location is small, the information about the overall environment is rigid and poor, and an egocentric reference system is used to decide the direction of movement. On the contrary, survey knowledge, for wayfinding, depends on the operator location awareness, which is generally considered an integrated form of representation with fast and route independent access to selected locations, dynamic and overviewing information of the environment layout, and structured in an allocentric coordinate system [18].

Even if both interfaces provide both kinds of spatial knowledge, the PDA could result in less access to survey knowledge, due to the need to switch screen and the time consumption required to retrieve, process, and design the map. In the desktop interface the survey and route knowledge are always and contemporary available on the screen. Consequently, we hypothesized a better operator performance driving the robot with the desktop, in comparison with the PDA interface, in the scenario conditions which require dynamic environment orienting abilities (i.e maze). Contrarily, no meaningful performance differences were expected in the navigation situations in which no survey information is required and in which information obtained in the route perspective would be sufficient to accomplish the task (i.e narrow spaces). Besides, we expected a better general performance for PDA users in the full visibility condition, as the operator has the possibility to see the robot either represented on the PDA display either in the real environment. This could plausibly decrease the information accessibility disparities between the two interfaces and take advantage of a more salient route information access deriving from a direct environment experience. In any case, we did not know if communications latency and lower computational power associated with the PDA device would significatively influence the performance. The experiments were made in order to validate these hypothesis, and to test if the mobility advantage related to a PDA could counterbalance the disadvantages associated with the device limitations.

4.4 Data analysis

We used ANOVA (Analysis of Variance) to analyze the data and consider only significative values (with a significance level set at 0,05). Roughly speaking, a population of samples is significant when their variations are imputable only to intrinsic factors, and not to casual ones. In order to accept a statistical significance, the p-value returned by ANOVA must be lower than the set level. Within ANOVA, the F-test is used to compare the total deviations of the two components, returning the ratio of the two variances (F-value or F), that is usually reported with the p-value.

First Experiment. For the exploration task analysis we have subdivided an exploration time of 10 minutes in twenty discrete values (from 0.5 to 10); then a 2x20 ANOVA on the explored area (in m^2) was carried on with the "between-participants" factor of Interface (Desktop and PDA) and the "within-participants" factor of Time (in

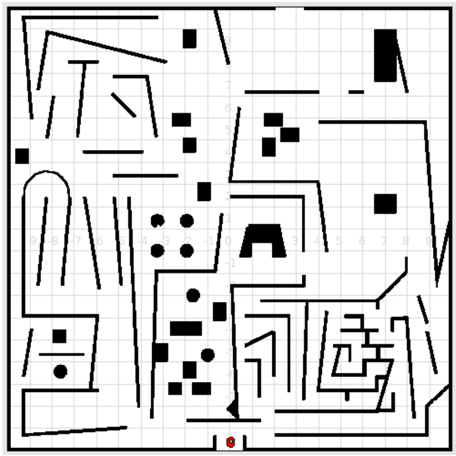


Figure 5. Virtual scenario for the first experiment

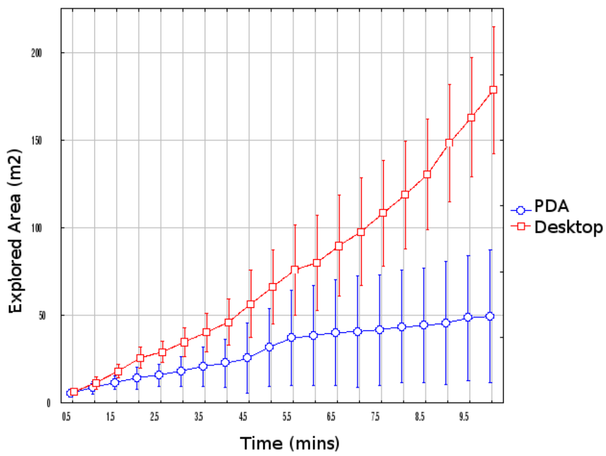


Figure 6. Covered area in square meters by the operator using the PDA (bottom curve) and the operator using the desktop interface

minutes from 0.5 to 10). The covered area was considered as a measure of the performance of the exploration.

Second Experiment. A one way ANOVA on navigation times was calculated to compare the interfaces in order to see if the differences under the PDA condition and the desktop condition had significant differences when the operator had to navigate without dealing with exploration (way finding).

Third Experiment. Two separated ANOVA's (since the visibility variable did not vary in the Desktop interface) on navigation times for each condition of PDA visibility: Total Visibility (TV) and Partial Visibility (PV), were carried out to study the PDA visibility effect on performance. For each of these analysis the design was a 2x3 with the Interface (Desktop and PDA) as a "between-participants" factor and Space Type (Maze, Narrow Spaces and Cluttered Area) as a "within-participants" factor. Three planned comparisons between Desktop and PDA for each space type were calculated afterwards to analyze the interface differences depending on the structural characteristics of the navigated space. Because the two analyzed populations can be characterized by a normal distribution, Student's t-test was used to verify a significant statistical difference between them..

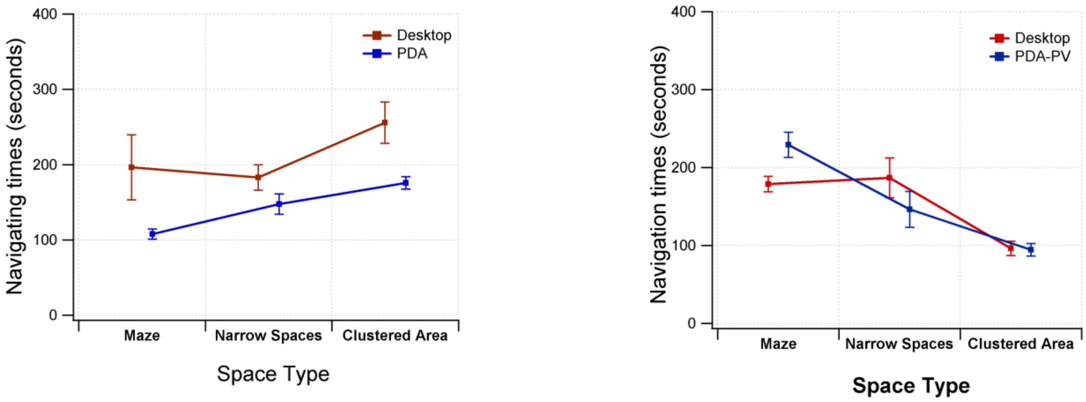
4.5 Results

First Experiment. Results are shown in Figure 6. A direct observation of the areas explored by the operator using the desktop and the operator using the PDA reveal that the former performs considerably better. The analysis showed a significant interaction between Interface and Time $[F(19, 361) = 13.65, p < .00001]$. A planned comparison for each level of time was calculated, indicating that at minute 1.5 of exploration the difference between Desktop and PDA, in terms of explored area, is just significant $[p < .05]$. Then it remains significant and grows at each level of Time.

Second Experiment. The first ANOVA on navigation times resulted non significant $[F < 1]$ revealing no difference among the driving times between interfaces.

Third Experiment. Results are shown in Figures 7(a) and 7(b). The plain observation of the figures indicate that the operator using the PDA with full visibility drove the robot faster, while in the partial visibility it depends on the kind of *Space Type*. To study if these differences were significant we ran the ANOVA tests.

The first 2x3 ANOVA with the "between-factor" *Interface* (Desktop and PDA-TV) and the "within-factor" of *Space Type* (Maze, Narrow Spaces and Cluttered Area) revealed a non significant interaction between them $[F(2, 32) = 1.43, p > .05]$; the main effects of *Interface* was instead significant $[F(1, 16) = 6.67, p < .05]$, revealing faster navigation times with the PDA interface in the total visibility condition in comparison with the desktop interface, independently from the *Space Type* (Figure 7(a)). The second analysis (partial visibility) showed a significant interaction between the *Interface* (Desktop and PDA-PV) and the *Space Type* $[F(2, 32) = 4.41, p < .05]$. Consequently three planned comparisons between the desktop interface and the PDA interface with partial visibility (Figure 7(b)) for each condition of *Space Type* were calculated and revealed that in the Maze condition, the desktop interface drives to faster navigation times in comparison with the PDA interface under partial visibility $[p < .05]$; no other significant differences were observed between the interfaces in the other *Space Type* conditions.



(a) Operator using PDA with full visibility of the scenario wrt. operator using desktop. Mean times and std. dev. are represented

(b) Operator using PDA with partial visibility of the scenario wrt. operator using desktop. Mean times and std. dev. are represented

From the t-test analysis a significant and a tendency to a significant difference between the desktop interface and the PDA interface was observed, respectively in the *Cluttered Area* $[p < .05]$ and *Narrow Spaces* $[p < .06]$. We hypothesize that collected data was not significant due to the reduc number of subjets.

5 Discussion

The results of the experiments clearly indicate a difference between interfaces depending on the type of task. In Table 1 the different cases that were considered in the nuclear plant scenario are illustrated, indicating which interface would perform better. The exploration task corresponds to the case in which operators must assess the state of and unknown area. It applies to situations after a disaster, in which he or she must drive the robot throughout the area looking for victims, damages, etc. Navigating, instead applies to the case in which the operator must follow a path in which he or she must not find a way (corridor, tunnel, etc.), more typical of maintenance operations. For analyzing the data we have considered that finding a way through the maze consists of an exploration task, driving alongside narrow spaces of a navigation task, while driving in the cluttered area is a combination of both.

Both interfaces are practically identical in the navigation task under the same condition of visibility (second experiment), a highly relevant distinction between them could be stated for the exploration task (first experiment). Here the results evidenced that at just 1.5 minute of exploration the desktop exploration area was greater than the PDA one; moreover this difference gradually increased with time. This result was predicted and analyzed in the hypothesis section. The data analysis of the third experiment shows that in total visibility condition the PDA interface results in a general better performance in terms of navigation times than the desktop interface independently from the space type. That is, the information the operator receives through the PDA, completed with the information he receives directly from the operating scenario, provides him or her with a better robot situational awareness (location and surrounding awareness [20]) for driving the robot. This implies that a PDA permits a successful task accomplishment when the robot is monitored by using both screen given information and real environment cues. This different kind of information integration together with the interface simplicity allows the operator to overcome the device limitations.

Exploration	Expl./Nav.	Navigation
Total Visibility		
PDA		
Partial Visibility		
Desktop	*Desktop/PDA*	*Desktop/PDA*
No Visibility		
Desktop		*Desktop/PDA*

Table 1. Best performance interface depending on the task and visibility

Concerning the partial visibility condition results indicate that an operator driving the robot with our desktop interface in a maze-like space, brings to faster navigating times than the operator using our PDA interface. We hypothesize that this effect is due to the amount of information given by the two interfaces: while indeed in the desktop local and global (survey perspective - location awareness) and tridimensional (route - surrounding awareness) perspectives are simultaneously available, in the PDA only one of these views is shown and the operator must switch between tabs to change it, employing more time. Even more, switching the tab forces adds an ulterior latency time due to the computational time to render the selected visualization mode. This occurs mostly in mazes, presumably because in these kind of environments a global configuration of the spatial structure (survey perspective) is needed in order to find a way out. This explanation is also supported by the t-test results, which indicate a general better performance of our PDA interface wrt. the desktop interface in cluttered and narrow spaces, which likely do not necessarily require survey knowledge to successfully navigate through.

We finally hypothesize that the differences between tasks could derive from their different information requests. It is presumable that all the information given by the desktop (local, global and tridimensional perspectives) is not necessary in the navigation task but indispensable in the exploration task, in order to give the required location awareness.

6 Conclusion and further work

In this paper we have studied the influence of operator mobility and task when controlling a robot using a PDA interface wrt. controlling a robot with a desktop interface. Even if the results here analyzed only apply to our interfaces, we believe that they can be generalized to the device and thus, our thesis is that similar results would be obtained if the same experiments were run with interfaces designed differently.

As a main conclusion we can state that the SA of the operator is smaller when using the PDA (mostly the location SA) due to the fact that the small size and low computation capacity of a PDA device does not allow to provide to the operator with the same amount of information that when using a desktop. *Nonetheless, the possibility of moving inside the operating scenario that permits a hand-held device can counterbarlance this disadvantage, as proved from the results*. In the future we will work in finding ways of enhancing the survey knowledge (location awareness) through the PDA interface, in order to diminish this differences. In any case, *results demonstrated that when the operator does not need to find a way (exploration) but to follow a path, both interfaces were feasible enough to drive the robot obtaining the same performance*.

Furthermore, considering that our anticipated work consists of a team of operators controlling a team of robots, besides providing robot situational awareness, we must study how to provide them with team situational awareness, in order to coordinate the team activities, transfer the control of the robots and *intelligently* allocate the control of the robots among the operators. We are planning for the next year to make experiments in which operators do not control the robot independently, but in which the operator using the PDA and the operator using the desktop, control simultaneously a team of robots.

ACKNOWLEDGEMENTS

We would like to thank the referees for their time and comments which helped to improve this paper.

REFERENCES

[1] Julie A. Adams, 'Critical considerations for human-robot interface development', Technical report, 2002 AAAI Fall Symposium: Human Robot Interaction Technical Report, (2002).

[2] G. Cohen, *Memory in the real world*, Hove: Erlbaum, 1989.

[3] Frauke Driewer, Markus Sauer, and Klaus Schilling, 'Design and evaluation of an user interface for the coordination of a group of mobile robots', in *17th Iternational Symposium on Robot and Human Interactive Communication. RO-MAN 2008*, pp. 237–242, (August 2008).

[4] Jill L. Drury, Brenden Keyes, and Holly A. Yanco, 'LASSOing HRI: analyzing situation awareness in map-centric and video-centric interfaces', in *Proceedings of the Second ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, pp. 279 – 286, (2007).

[5] Jill L. Drury, Holly A. Yanco, and Jean C. Scholtz, 'Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition', *Human-Computer Interaction Journal*, (January 2004).

[6] Mica R. Endsley, 'Design and evaluation for situation awareness enhancement', in *Proceedings of the Human Factors Society 32nd Annual Meeting*, pp. 97–101, (1988).

[7] Terrence Fong, Charles E. Thorpe, and Charles Baur, 'Advanced interfaces for vehicle teleoperation: Collaborative control, sensor fusion displays, and remote driving tools', *Autonomous Robots*, **11**(1), 77–85, (2001).

[8] Brian P. Gerkey, Richard T. Vaughan, and Andrew Howard, 'The player/stage project: Tools for multi-robot and distributed sensor systems', in *11th International Conference on Advanced Robotics (ICAR 2003), Portugal*, pp. 317–323, (June 2003).

[9] Th. Herrmann, 'Blickpunkte und blickpunktsequenzen.', in *Sprache & Kognition*, volume 15 of *217-233*, (1996).

[10] Hande Kaymaz-Keskinpala, Kazuhico Kawamura, and Julie A. Adams, 'Pda-based human-robotic interface', in *Proceedings of the IEEE International Conference on Systems, Man & Cybernetics*, volume 4, pp. 3931 – 3936, (2003).

[11] Amir Naghsh, Jeremi Gancet, and Andry Tanoto, 'Analysis and design of human-robot swarm interacion in firefighting operations', in *17th Iternational Symposium on Robot and Human Interactive Communication. RO-MAN 2008*, pp. 255–260, (August 2008).

[12] Curtis W. Nielsen and Michael A. Goodrich, 'Comparing the usefulness of video and map information in navigation tasks', in *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, HRI 2006*, pp. 95–101, (2006).

[13] Dan R. Olsen and Michael A. Goodrich, 'Metrics for evaluating human-robot interactions', in *Proceedings of PERMIS 2003*, (2003).

[14] Jean Scholtz, Jeff Young, Jill L. Drury, and Holly A. Yanco, 'Evaluation of human-robot interaction awareness in search and rescue', in *Robotics and Automation, 2004. Proceedings ICRA'04*, volume 3, pp. 2327 – 2332. IEEE, (May 2004).

[15] A.W. Siegel and S.H. White, 'The development of spatial representations of large-scale environments', in *Advances in Child Development and Behavior, NewYork*, ed., H.W Reese, volume 10, pp. 9–55. Academic Press, (1975).

[16] Barbara Tversky, 'Spatial mental models', in *The Psychology of Learning and Motivation, San Diego*, ed., G.H Bower, pp. 109–145, (1991).

[17] Alberto Valero, Fernando Matia, Massimo Mecella, and Daniele Nardi, 'Pro-active interaction for semi-autonomous mobile robots. introducing the task allocation service', in *IEEE ICRA08 Workshop: New Vistas and Challenges in Telerobotics, California*, (2008).

[18] Steffen Werner, Bernd Krieg-Brückner, Hanspeter A. Mallot, Karin Schweizer, and Christian Freksa, 'Spatial cognition: The role of landmark, route, and survey knowledge in human and robot navigation', in *GI Jahrestagung*, pp. 41–50, (1997).

[19] Holly A. Yanco and Jill L. Drury, '"where am i?" acquiring situation awareness using a remote robot platform', in *Proceedings of the IEEE International Conference on Systems, Man & Cybernetics: The Hague, Netherlands, 10-13 October 2004*, pp. 2835–2840, (2004).

[20] Holly A. Yanco, Jill L. Drury, and Jean Scholtz, 'Awareness in human-robot interactions', in *Proceedings of the IEEE Conference on Systems, Man and Cybernetics, Washington, DC, October 2003*, (2003).

Proceedings New Frontiers in Human-Robot Interaction, K. Dautenhahn (Ed.), symposium at the AISB09 convention, 8-9 April, Edinburgh, Scotland

Baby steps: A design proposal for more believable motion in an infant-sized android

Silpa Wairatpanij, Himalaya Patel, Gary D. Cravens, and Karl F. MacDorman[1]

Abstract. In studying human interaction, an android can serve as a precisely controlled apparatus to elicit human response. However, if an android is to substitute for a human being in social, cognitive, and neuroscientific experiments, it is essential to control for the effects of appearance by designing it to look and move as much like a human being as possible. The goal of this project is to create an infant-sized android to use in experiments. We propose design concepts for android joints and joint control to imitate human muscle control and appearance. The synchronized motion of servomotors and anatomically correct muscle-joint relations combine with the passive motion of elastic muscles to control the android's movements. The implementation of an android leg exhibits these concepts.

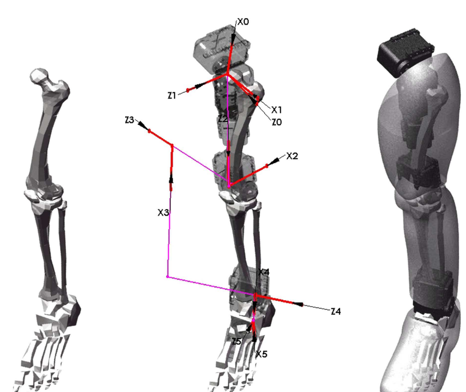
1 INTRODUCTION

People can perceive and respond to very humanlike robots (androids) as if they were human [18]. This cannot be said of less humanlike robots, because they fail to control for such extraneous variables as the effects of their nonhuman appearance. Thus, androids open the way for a new methodology to explore human cognition and interaction. They can be used in place of human stimuli as an experimental apparatus in social, cognitive, and neuroscientific experiments. This has the advantage of increasing experimental control, because androids—unlike human actors—can be programmed to respond with consistency and precision. In addition, an android's physical embodiment affords a sense of physical presence that human interaction with other technological alternatives lacks, such as interaction with computer-generated characters or recorded videos. An android's heightened ecological validity as compared with other media is of particular value during interactions with infants, because a baby cannot be told what to do during an experiment.

While it has long been supposed that infants develop cognitively by imitating adults [20], researchers in infant development have found evidence that infants first begin learning by being imitated by adults [13]. Contingency is a key factor in the infant's learning process. Contingency denotes the property of one event depending on another. For example, if one robot beeps when an infant vocalizes and another beeps randomly, an infant will make vocalizations (probes) to determine rapidly which robot is responding contingently. Movellan [23] has argued that infants are nearly optimal contingency detectors endowed with a rapid learning mechanism to maximize the amount of information they acquire about those in their surroundings. Adult imitation of the infant provides contingent feedback about what the infant's body is doing. The infant starts by reproducing the effects of the adult's actions and then learns to reproduce the actions themselves [13]. Associative learning enables observed contingencies to shape future interactions and transform agency [5]. Contingent feedback also enables the development of mirror neurons in the infant's brain. Mirror neurons fire both when someone performs an action and when that person sees someone else perform the same action. Thus, they enable us to understand the intention behind another person's action by putting ourselves in that person's place.

Our goal is to develop a baby android to study how contingency and timing shape infant interactions and their role in learning nonverbal behavior. The investigators will use the baby android for the fine-grained study and analysis of contingency in both infant-android interaction and adult-android interaction. It has been established that even three-month-old infants are highly sensitive to contingent feedback, reacting negatively when a live interaction with their mothers through a video link was replaced with a noncontingent, recorded interaction [24]. We plan to explore contingency by interacting with infant participants through an android that is controlled by telepresence. This method, commonly referred to as *Wizard-of-Oz,* entails a person responding to information from the robot's sensors and controlling its actuators through haptic devices like a joystick or a Wii remote. The influence of timing on contingency will be studied by inserting delays of varying duration into these technology-mediated interactions. Machine learning and data mining techniques will be used to analyze the statistical relation between the responses of interaction partners by using motion capture equipment.

In using a robot to investigate the micro-dynamics of infant interaction, it is important to control for the effects of the robot's appearance. If an infant responded differently to a mechanical-looking robot than to another infant, it would be difficult to determine whether the



1A 1B 1C

Figure 1. Design Concept for Infant Leg

[1] Indiana University, USA, MacDorman.com

Table 1. Human synovial joint types [10]

Joint Type	Motion	Locations	Motors
Hinge	Uniaxial	Ankle, elbow, interphalangeal	1
Pivot	Uniaxial	Vertebral column	1
Saddle	Biaxial	Hand, foot	2
Condyloid	Biaxial	Wrist	2
Ball and socket	Multiaxial	Hip, shoulder	3
Plane	Gliding	Scapula	N/A

robot's appearance, motion quality, or contingency were the cause. Of particular concern with robots that are intended to look human is the uncanny valley [18]. The negative evaluations associated with the uncanny valley can result from a mismatch among robot elements, such as a robot that has human-looking eyes but metallic "skin" [12]. However, even in a consistently human-looking robot, negative evaluations can result when the robot's kinematics and dynamics are unable to match its morphological realism. For example, covering a mechanical-looking robot (e.g., Honda's Asimo) with a human-looking "skin" might seem eerie to observers, especially when it starts to move because of the placement of its joints, their movement beyond allowable human ranges, uneven stretching of the skin, absence of apparent muscle movement, lack of inertial dampening, fixed compliance, and so on.

This paper proposes design concepts for android joints and applies these concepts to building an infant-sized android to be used in the aforementioned experiments. Section 2 compares human and robot joint systems. Section 3 explains the proposed android design concepts. Section 4 presents the results of the concept simulation. Section 5 explains our directions for future work.

2 HUMAN VS. ROBOT JOINT CONFIGURATION

The muscles and skeletal system of the human body work together under the control of the nervous system to maintain posture and produce movement [9]. The muscles are arranged into antagonistic sets which control motion around a joint. A given joint may have from one to three axes of rotation and sets of antagonistic muscles (Table 1). Contracting a set of muscles applies a torque at the joint. The net movement of the joint results from the net torque of opposing sets of muscles around the axis. The nervous system allows one set of muscles to relax while contracting the opposing set. When a muscle contracts, the cross-sectional area of the muscle increases, causing the skin around the muscle to bulge. When the muscle relaxes, the skin flattens. For example, flexion (bending) and extension (straightening) of the knee involve movements in opposite directions: The biceps femoris, gastrocnemius, soleus control flexion, and the quadriceps femoris controls extension. The nervous system is continuously sending signals to antagonistic sets of muscles to slightly adjust their contraction for the body to maintain its posture. The nervous system continually adjusts the body's posture by processing input from proprioception (position sensors in the body), comparing the present state to the desired posture, and moving opposing muscle sets to maintain the desired posture.

Unlike human limbs, robot links are moved by actuators-at the actuator axis for revolute actuators or at the pivot point for prismatic actuators. Although the properties of revolute actuators differ from the human muscle system, they are widely used because of their compact size, high precision, and low maintenance. Servomotors have been mainly used in humanoid robots including Asimo [11], Hubo [26], HRP [15], HRP-2 [14], iCub [31], Qrio, SDR-4X, Kenta [21], and Kenji [22]. Many robotics researchers conceive of robots as tools helping people performing household tasks [15, 14, 21, 22, 29, 30]. Thus, motors and joints in humanoid robots are typically placed to yield high motor performance. In addition, they are typically aligned perpendicularly or in parallel to simplify the kinematic chain calculation. Human beings in natural postures do not have any two joints that are perpendicular or parallel. The characteristics of the actuators and the kinematic chain design of humanoid robots results in very robot-like movement. However, observers do not feel perturbed by the unnatural movement, because the humanoid robot does not look human [4].

Much research has attempted to create biologically inspired robots using prismatic actuators, which have contraction and extension properties similar to human muscle. Pneumatic actuators were used to move such human replicas as Repliee Q1 and Q2 [19] and Geminoid [25]. The androids were covered by clothes, so the skin movement was not noticeable. Joint configuration using pneumatic or hydraulic actuators can be similar to the joint configuration of humans. However, without an antagonistic arrangement of actuators, the limbs of the robots yielded too much compliance and could overshoot the intended target at higher speeds. Their pneumatic actuators also need more maintenance than servomotors and require a large, external air compressor. Many attempts have been made to create biomimetic robot joints [21, 22, 30, 3, 16]. One of them used electroactive polymer (EAP) actuators in an antagonist setup around the pivot point of the robot's eyeballs [3]. Even though the system can emulate human movement, EAP actuators have been capable of only a small displacement [16, 2], which is insufficient to move robot limbs, although EAP actuators capable of larger displacements are under development [1]. However, these systems are good at making facial expressions [26]. Festo AG has created Airic's_arm using a proprietary pneumatic system, which employs contraction membranes without piston rods [8]. The arm features a human skeletal structure, including shoulder blade movement. However, the system also requires an external air compressor, and the cost is still too high for mass production.

We believe a carefully constructed actuation model underneath the skin can greatly enhance the human motion realism of an android robot. We wish to find a simple way to achieve natural movement while maintaining low costs and a compact design in an android replica of a 13-month-old baby. The size of the actuators and the placement in the android body were the determining factors for the selection of the age of the android. We found that 13 months was the minimum age at which the body size could accommodate all the actuators.

3 PROPOSED CONCEPTS

A robot's kinematics constrains the appearance of its movement. Therefore, the mechanics of its underlying structure should be as close to that of a human being as existing technology allows. Three main concepts are described in this section: the robot's internal construction, its controlling schema, and its external construction.

3.1 The Selection of Actuators

A mechanical actuator that has the physical properties of human muscles does not exist. Prismatic actuators are the closest. However, they do not satisfy the requirements of this project because of their large size and high maintenance costs. The actuators should be small enough to fit within the proportions of an infant, and their operation should not interfere with apparent muscular movement, as described in section 3.3. For these reasons, we use revolute actuators. Various combinations of revolute actuators can mimic human joint motions, as shown in the last column of Table 1, except for gliding joints, as indicated in last entry of that column. Because the goal of the design is human-looking movement, the placement of each actuator axis has to align closely with the corresponding human joint axis.

3.2 Joint Control

Robot movements often look awkward because joints are moving independently of each other; furthermore, the motion of the joints can include abrupt changes in acceleration. The perceived visual salience of independent joint movement are highlighted when human dancers perform moves in the "robot" dance style. Under normal conditions, human muscles work differently. Most influence more than one skeletal joint [10]. Moreover, kinematic calculations have focused on the input angles of joints to accomplish the position and orientation of the end effectors, for example, the hands and feet [11, 26, 15, 14]. Without paying close attention to the angle outcomes of intermediate joints, the inverse kinematics calculation might result in limb positions that are abnormal or impossible for human beings—a disturbing sight to watch. In computer graphics (CG), animators face difficulty in moving a character's limbs naturally by giving an angle input to each joint. Part of the problem in mimicking human motion is that some human muscles simultaneously apply torque to two or more joints or axes of a single joint.

To overcome this problem, the controlling schema should involve muscle-like control of the joints rather than joint angle control. As with the biological model, muscle movements should determine limb angles. Movement of a given joint may result in movement of the adjacent joints. Moving one joint but failing to mimic the effects of the controlling muscles on other joints can cause the movement to look eerie.

3.3 Mimicking Muscle Shape

In most cultures and climates, some degree of skin exposure is typical and, indeed, beneficial to communication; clothing can obscure gestures, facial expressions, and other nonverbal behavior. Because the surface of the android is visible where clothing is absent, care must be taken to make it look human. However, simply covering moving mechanic parts with synthetic skin might create a strange twisting and stretching effect. In addition, the selected actuators lack the shape changing properties of human muscles. To create a believable muscle effect, a layer of muscle-shape rubber material should be inserted between the actuator housing and the synthetic skin. This has the following advantages: First, when rubber muscles are stretched or contracted, they help distribute the stretching of the android skin more evenly. Second, the cross-sectional area of rubber muscle decreases when stretched and increases when contracted, which is the same as human muscle.

4 IMPLEMENTATION

Although a different part of the body could have been chosen for the initial work, the leg was chosen because of its relative simplicity compared to other parts of the body. For example, it is reasonable to ignore the motion of the toes for this model because of their limited size and range of motion; however, ignoring the motion of the fingers would not be an option when designing the arm.

One approach that would not work for this project is to have a simulation movie of the leg by itself (or any other part of the body) to get feedback from individuals as to whether the motion appears natural. The reason for this is that the leg is not an independent system; motion at the hip partially depends on the hip's connections to the torso. One of the fundamental approaches of this work is to take into account the lack of independence of motion of the joints. In this section the concept of the infant android leg is implemented on a CAD program for engineering.

4.1 Actuators and Their Placement

Robotis Dynamixel AX-12 servomotors [28] were selected as the android's actuators for the following reasons. First, the servomotors are small enough to fit in infant limbs. Second, small servomotors require no maintenance once installed. Third, each of these particular servomotors can be controlled in velocity mode, imputing the desired angular velocity and clockwise or counter-clockwise direction. Changes in angular velocity can be translated into torque. Fourth, the actuators have enough torque to lift up each limb. In the planned experiments the robot is not required to walk but is seated on a chair, table, or floor. Thus, one actuator does not need to hold the entire weight of the robot.

Table 2. Joint Movement of the Lower Limb

			Maximum Values	
Joint Type	Actuator	Movement	Displ. $(deg.)$	Accel. $(\frac{deg.}{s^2})$
Hip	A	Flexion	120	660
Ball and		Extension	-45	
socket	B	Adduction	20	280
		Abduction	-50	
	C	Lateral rotation	80	400
		Medial rotation	-20	
Knee	D	Flexion	160	648
Hinge		Extension	-2	
Ankle	E	Dorsiflexion	45	300
Hinge		Plantarflexion	-30	
Foot	F	Inversion	35	200
Pivot		Eversion	-15	

A computer graphics model of the infant skeleton was created to verify the accuracy of the actuator placements (Figure 1A). Actuators, then, were placed to ensure the alignment of each actuator axis with the skeletal joint axis. Figure 1B shows the result of the placements. The types of lower limb joints, listed in Table 1, are ball-and-socket joint at the hip, hinge joint at the knee, hinge joint at the ankle, and many small saddle joints at the ankle. However, the saddle joints were simulated by a single pivot joint. There are a total of 12 major movements of a human leg with 2 antagonistic movements per axis. Thus, the robot requires a total of six actuators for each leg. Table 2 describes the skeletal joints of the leg, joint movement, and actuators involved in each movement. Movements are illustrated in Figure 1B. Some actuator placements were shifted owing to a lack of space inside the robot's limbs. Nevertheless, they still maintain the projected axis of the skeletal joint axis.

Table 3. Link Parameters

i (actuator)	a_{i-1}	α_{i-1}	d_i
0 (A)	0	0	0
1 (B)	−90.00	0	0
2 (C)	−87.06	0	−143.44
3 (D)	−86.33	5.00	−179.19
4 (E)	149.60	144.36	92.90
5 (F)	80.83	27.95	0

The resulting placement of the actuators clearly contrasts with that of humanoid robots. For example, at the anatomically normal human standing position, the knee axis is not parallel to the ankle axis, but tilted inwards about ten degrees. Table 3 shows link parameters in the Denavit-Hartenberg notation derived from actuator placement in Figure 1B.

All the actuators are covered by an inner shell: fabricated plastic cases holding the servomotors in place and smoothing out their sharp edges. The shape of the inner shell is similar to the shape of a limb, as shown in Figure 1C. The outside of the shell was designed to be smooth with no sharp edges to impede the movement of the muscle layer.

4.2 Elastic Muscles Mimicking the Changing Shape of Moving Muscles

A rubber sheet was designed to be cut in a narrow leaf shape in a variety of lengths and widths depending on the structure of each muscle. Each end of a rubber muscle will be attached to the inner shell of the leg.

For example, the *tibialis anterior* contributes to inversion, tilting the foot towards the center of the body, and dorsiflexion, rotating foot at the ankle upward. Thus, servo F will be turned 35 degrees inward, the maximum displacement of inversion, and servo E will be turned 45 degrees upward, the maximum displacement of dorsiflexion. This position ensures the minimum stretched displacement of the elastic material corresponding to the *tibialis anterior* as allowed by ankle motion. This results in the maximum allowed cross-sectional area of the elastic material for that muscle. Then, one end of the muscle will be glued to the under surface toward the inside of the foot and the other end will be glued to the front of the calf below the knee. Toward the completion of the android, all limbs will be covered with rubber skin material.

4.3 Control System

Out of more than 50 muscles in the human lower limb [10], we selected only 12 muscles that related to major movements of the leg. Our selection criterion was the muscle's contribution to a noticeable movement. An infant's subcutaneous layer of fat is relatively wide compared to healthy adults. This is partly caused by the underdeveloped state of an infant's muscles. It is also caused by the infant's higher surface area-to-volume ratio. This results in heat loss from a skin surface area that is large relative to the volume of the body generating heat through metabolism. Therefore, infants have a relatively large fat layer as a form of insulation [17]. So the movement of infant muscles is not as visible as in most adults.

For the purposes of our design, *gluteus maximus* contributes 80% of the total torque for extension at the hip joint with the remaining 20% provided by *biceps femoris* and muscle group B. The *gluteus maximus* also provides 70% of the torque for lateral rotation at the hip joint with the remaining 30% provided by *quadriceps femoris, biceps femoris,* and muscle group B. Minor contributions of muscles to movement around a joint axis were ignored. For example, *gluteus maximus* is also involved in flexion at the hip joint, but the contribution is negligible and therefore excluded.

Lacking data regarding the contribution of various muscles to torque, we assumed that the maximum torques of the antagonistic muscles or muscle groups are equal at each axis. Thus, a joint stops moving or moves at constant speed when the antagonistic muscles simultaneously contract by an equal amount.

Let A be a 6×12 rectangular matrix containing the percentage of torque each muscle contributes to each movement. Let $\hat{x}$ be a 12×1 input vector containing for each muscle the percentage of its maximum torque to be applied for the duration of the movement frame (system interval). And, let $\hat{b}$ be a 6×1 vector.

$$\frac{A}{100} \cdot \hat{x} = \hat{b} \quad (1)$$

From the result of (1), $\hat{b}$ constitutes the directions and net torques at the axes. In angular motion,

$$\tau = I \cdot \alpha \quad (2)$$

where τ is torque, I is the moment of inertia, and α is the angular acceleration. To simplify the control system calculation, we let I be a constant denoted by k because of its small mass and small distance from the axis of rotation to the center of gravity of the limb.

Another simplification in this system is the assumption of linearity. The muscle itself is not a linear system; however, typical length-tension curves for skeletal muscle comprise the sum of active and passive tension components. Although the sum of the tension components is not linear, its deviation from linear for most of the range of motion is limited enough that we feel that the assumption is justified [27]. Thus,

$$\tau_{max} = k \cdot \alpha_{max} \quad (3)$$

From (3), angular acceleration is scaled by the net torque applied to a joint. To find an order of magnitude estimate of α_{max} for each axis, we assumed that the full range of movement can be completed in approximately one second with constant angular acceleration for the first half of the movement and constant angular deceleration for the second half of the movement. The extreme case occurs when a joint is moved in a full range of motion. Under such a circumstance, each muscle related to the movement direction contracts with maximum torque to reach the maximum angular acceleration. We can obtain an order-of-magnitude estimate for maximum acceleration as shown in the last column of Table 2.

Two variables that the servomotor controller requires for each motor are the goal position and the desired angular velocity. The desired torque can be computed by these two variables. First, a comparison of the current position and the goal position indicates the turning direction—maximum displacement of each movement, from Table 2, was used for the goal position. Second, the desired torque is calculated from the angular acceleration—the angular acceleration is the change of the angular velocity with respect to time. Thus, in each time period of the trajectory calculation, the angular acceleration can be calculated by the following equation:

$$\alpha = \frac{d\omega}{dt} \quad (4)$$

where ω is the angular velocity.

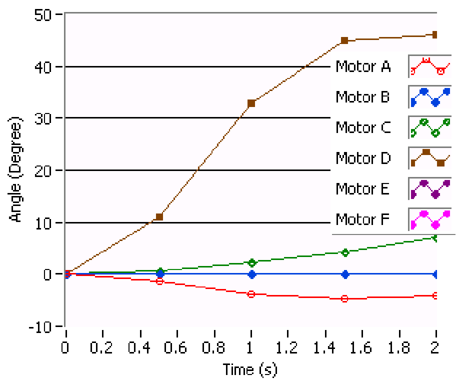


Figure 2. The angular motion of motor A to F in two seconds

$$\alpha = \frac{\omega_2 - \omega_1}{\Delta t} \quad (5)$$

where ω_1 is the initial angular velocity of the time step and ω_2 is the final angular velocity; Δt is the system frame interval. Thus,

$$\omega_2 = \omega_1 + a \cdot \Delta t \quad (6)$$

This equation can be rewritten in term of percentage of maximum angular acceleration as

$$\omega_2 = \omega_1 + \%\alpha_{max} \cdot \Delta t \quad (7)$$

And, θ can be found by $\theta = \int_{t_1}^{t_2} \omega \, dt$.

Table 4. Net Torque Input to Muscle at Four Time Frames

Muscle	$\hat{x}_{t1}$	$\hat{x}_{t2}$	$\hat{x}_{t3}$	$\hat{x}_{t4}$
Tibialis anterior	0	0	0	0
Fibularis	0	0	0	0
Tibialis posterior	0	0	0	0
Gastrocnemius & Soleus	0	0	0	0
Quadriceps femoris	0	0	12%	1.50%
Biceps femoris	15%	0	0	0
Iliopsoas	0	0	0	0
Gluteus maximus	0	0	0	0
Group A	0	0	0	0
Group B	0	0	0	0
Group C	0	0	0	0
Gracilis	0	0	0	0

For example, Table 4 shows the vector $\hat{x}$, input of the system, in four time frames, from t_1 to t_4, to flex the knee joint at about 45 degrees in two seconds. The system interval, Δt, is half a second. *Biceps femoris* contracts 15 percent in the first interval.

Quadriceps femoris then contracts at 12 and 1.5 percent to counteract the motion of the flexing knee in the third and fourth interval. Figure 2 shows the trajectory simulation results of the input from Table 4. All servomotors start at 0 degree when $t = 0$. The servomotor at the knee joint, motor D, decelerates when *quadriceps femoris* contracts at $t = 1$. The results clearly show the effect of contracting biceps femoris on the hip joint, motor A, and motor C (Fig. 3).

5 FUTURE WORK

5.1 Telepresence Data Mining for Contextualized Interaction

A major limitation of social robots that support human interaction is that the interaction is designed in advance by robot engineers. However, it is difficult to plan for the social context, because it is not a static property of situations. Instead, it is created dynamically during the course of the interaction [6]. Interactional microdynamics are not under conscious control. As a result, they are often overlooked by engineers who do not build them into the design specification. For this reason, the robots fail to elicit our sense of human presence, which is the most highly therapeutic aspect of the interaction.

We propose an entirely new methodology for developing an android that can support social interaction. Instead of "designing the interaction," the engineer builds a system in which a human interacts with another human by controlling the android. Furthermore, the human can be limited to observing the same processed information from the android's sensors that an android control system would. After collecting a large amount of data from these robot-mediated interactions, data mining and machine learning techniques are applied to extract patterns of interaction that are then implemented in the robot so that it can function autonomously.

5.2 Potential Applications for Eldercare

Socially-assistive robots have been successfully employed for companionship, social mediation, monitoring care, and encouragement in performing rehabilitation exercises [7]. Just as robots have served as rehabilitation coaches for patients recovering from heart surgery or stroke, they have the potential to motivate older adults to exercise more to reduce obesity and improve cardiovascular health to prevent these ailments. Three major threats to quality of life in older adults are delirium, dementia, and depression, which is often associated with loneliness owing to social isolation. Robot pets, such as Paro, have been used successfully in nursing homes in Japan, the United States, and Europe, for companionship and as a vehicle to stimulate social interaction among patients [32]. Although the cognitive capacities of these robots are extremely limited compared to those of people, animal pets, and other robots, nonverbal cues such as head tracking and touch response can create a sense of presence that alleviates loneliness and stimulates sharing.

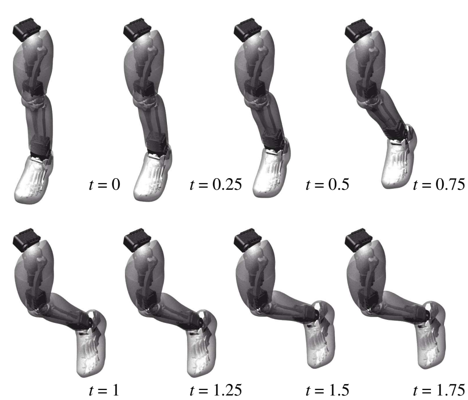


Figure 3. Two seconds of the simulated motion

Simple toys like Tamagochi and Amazing Amanda demonstrate that a device requiring human nurturance can be a "killer application" with vast market potential. A more complex android that combines humanlike nonverbal response with an infant-like form has the potential to have a far greater therapeutic impact in alleviating loneliness than these simple devices. The android baby will succeed to the extent that it is able to "press our Darwinian buttons" by mimicking subconscious behavior that elicits human response [32]. This information tells us, "Somebody's there," which can provide comfort. We propose to extend this by developing simple verbal communication and shared engagement in activities, such as exercise and games. This method has been successfully used by robots like Robovie in primary schools in Japan.

6 CONCLUSION

We have proposed a design for realistic joint movements and illustrated the underlying concepts by applying them to a planned implementation: the leg of a baby-sized android. There is more room for improvement regarding the complexity of the input script to yield a desired posture, for example, by implementing a neural network control scheme. We intend to apply these concepts throughout the body design before fabricating android parts. This project is still in development. The completed android hardware will be controlled mainly by scripted movements or by an experimenter using the Wizard-of-Oz method.

ACKNOWLEDGEMENTS

This research is supported by an IUPUI Signature Center grant.

REFERENCES

[1] Y. Bar-Cohen, K. J. Kim, H. R. Choi, and J. D. W. Madden, 'Electroactive polymer materials', *Smart Materials and Structures*, **16**(2), (2007).

[2] Y. Bar-Cohen, S. Leary, A. Yavrouian, K. Oguro, S. T. Tadokoro, J. Harrison, J. Smith, and J. Su, 'Challenges to the transition of IPMC artificial muscle actuators to practical applications', in *Material Research Society Symposium Proceedings Volume 600: Electroactive Polymers (EAP)*, eds., Q. M. Zhang, T. Furukawa, Y. Bar-Cohen, and J. Scheinbeim, Warrendale, PA, (1999). Materials Research Society.

[3] F. Carpi and D. De Rossi, 'Bioinspired actuation of the eyeballs of an android robotic face: Concept and preliminary investigations', *Bioinspiration and Biomimetics*, **2**(2), S50–S63, (2007).

[4] T. Chaminade, J. Hodgins, and M. Kawato, 'Anthropomorphism influences perception of computer-animated characters' actions', *Social Cognitive and Affective Neuroscience*, **2**(3), 206–216, (2007).

[5] S. J. Cowley, 'How human infants deal with symbol grounding', *Interaction Studies*, **8**(1), 83–104, (2007).

[6] P. Dourish, 'What we talk about when we talk about context', *Personal and Ubiquitous Computing*, **8**(1), 19–30, (2004).

[7] Feil-Seifer, K. Skinner, and M. J. Mataric, 'Benchmarks for evaluating socially assistive robotics', *Interaction Studies*, **8**(3), 423–439, (2007).

[8] Festo AG and Co. KG. Airic's_arm. World Wide Web. http://www.festo.com/INetDomino/coorp_sites/en/ffeed49f2394ea43c12572b9006f7032.htm.

[9] J. T. Hansen and B. M. Koeppen, *Netter's Atlas of Human Physiology*, Icon Learning Systems, Teterboro, NJ, 1st edn., 2002.

[10] J. T. Hansen and D. R. Lambert, *Netter's Clinical Anatomy*, Icon Learning Systems, Teterboro, NJ, 1st edn., 2005.

[11] M. Hirose, Y. Haikawa, T. Takenaka, and K. Hirai, 'Development of humanoid robot ASIMO', in *IROS 2001: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Workshop 2*, (2001).

[12] C.-C. Ho, K. F. MacDorman, and Z. A. D. Pramono, 'Human emotion and the uncanny valley: A GLM, MDS, and ISOMAP analysis of robot video ratings', in *HRI '08: Proceedings of the Third ACM/IEEE International Conference on Human-Robot Interaction*, pp. 169–176, New York, (2008). ACM Press.

[13] S. S. Jones, 'Imitation in infancy: The development of mimicry', *Psychological Science*, **18**(7), 593–599, (2007).

[14] K. Kaneko, F. Kanehiro, S. Kajita, H. Hirukawa, T. Kawasaki, M. Hirata, K. Akachi, and T. Isozumi, 'Humanoid robot HRP-2', in *ICRA 2004: Proceedings of the 2004 IEEE International Conference on Robotics and Automation*, volume 2, pp. 1083–1090, (2004).

[15] K. Kaneko, F. Kanehiro, S. Kajita, K. Yokoyama, K. Akachi, T. Kawasaki, S. Ota, and T. Isozumi, 'Design of prototype humanoid robotics platform for HRP', in *IROS 2002: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2431–2436, (2002).

[16] R. Kornbluh, R. Pelrine, J. Eckerle, and J. Joseph, 'Electrostrictive polymer artificial muscle actuators', in *Proceedings of the 1998 IEEE/RSJ International Conference on Robotics and Automation*, volume 3, pp. 2147–2154, (1998).

[17] G. Lowrey, *Growth and Development of Children*, Year Book Medical Publishers, Chicago, 8th edn., 1986.

[18] K. F. MacDorman and H. Ishiguro, 'The uncanny advantage of using androids in social and cognitive science research', *Interaction Studies*, **7**(3), 297–337, (2006).

[19] D. Matsui, T. Minato, K. F. MacDorman, and H. Ishiguro, 'Generating natural motion in an android by mapping human motion', in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, (2005).

[20] A. Meltzoff, 'The human infant as imitative generalist: A 20-year progress report on infant imitation with implications for comparative psychology', in *Social Learning in Animals: The Roots of Culture*, eds., B. Galef and C. Heyes, 347–370, Academic Press, New York, (1996).

[21] I. Mizuuchi, R. Tajima, T. Yoshikai, D. Sato, K. Nagashima, M. Inaba, Y. Kuniyoshi, and H. Inoue, 'The design and control of the flexible spine of a fully tendon-driven humanoid "Kenta"', in *IROS 2002: Proceedings of the 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 3, pp. 2527–2532, (2002).

[22] I. Mizuuchi, T. Yoshikai, Y. Nakanishi, and M. Inaba, 'A reinforceable-muscle flexible-spine humanoid "Kenji"', in *IROS 2005: Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4143–4148, (2005).

[23] J. R. Movellan, 'An Infomax controller for real-time detection of contingency', in *Proceedings of the International Conference on Development and Learning*, (2005).

[24] L. Murray and C. Trevarthen, 'Emotional regulation of interactions between two-month-olds and their mothers', in *Social Perception in Infants*, eds., Tiffany M. Field and Nathan A. Fox, Ablex Publishing, Norwood, NJ, (1985).

[25] Shuichi Nishio, Hiroshi Ishiguro, and Norihiro Hagita, 'Can a telep-operated android represent personal presence? A case study with children', *Psychologia*, **50**(4), 330–342, (2007).

[26] J.-H. Oh, D. Hanson, W.-S. Kim, I. Y. Han, J.-Y. Kim, and I.-W. Park, 'Design of android type humanoid robot albert HUBO', in *IROS 2006: Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1428–1433, (2006).

[27] R. A. Rhoades and D. R. Bell, *Medical physiology: Principles for clinical medicine*, Lippincott Williams and Wilkins, Philadelphia, 3rd edn., 2009.

[28] Robotis. Robot exclusive actuator dynamixel. World Wide Web. http://www.robotis.com/zbxe/dynamixel_en.

[29] S. Roccella, M. C. Carrozza, G. Cappiello, P. Dario, J. J. Cabibihan, M. Zecca, H. Miwa, K. Itoh, and M. Marsumoto, 'Design, fabrication and preliminary results of a novel anthropomorphic hand for humanoid robotics: RCH-1', in *IROS 2004: Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 1, pp. 266–271, (2004).

[30] Y. Sodeyama, I. Mizuuchi, T. Yoshikai, Y. Nakanishi, and M. Inaba, 'A shoulder structure of muscle-driven humanoid with shoulder blades', in *IROS 2005: Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4028–4033, (2005).

[31] N. G. Tsagarakis, F. Becchi, L. Righetti, A. J. Ijspeert, and D. G. Caldwell, 'Lower body realization of the baby humanoid - "iCub"', in *IROS 2007: Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3616–3622, (2007).

[32] S. Turkle, 'Authenticity in the age of digital companions', *Interaction Studies*, **8**(3), 501–517, (2007).

Preferences and Perceptions of Robot Appearance and Embodiment in Human-Robot Interaction Trials.[1]

Michael L. Walters, Kheng Lee Koay, Dag Sverre Syrdal, Kerstin Dautenhahn and René te Boekhorst.[2]

Abstract. Outcomes are presented from experiments on the effect of participants' individual *preferences* for robot appearance and height on their preferences towards and perceptions of live robots. Participants who expressed a preference for a mechanical looking robot, tended to prefer all robot types to stay further away than those participants who expressed a preference for more humanoid robots. A majority group of two thirds (68.5%) preferred a robot which they personally perceived as having an extrovert and agreeable personality and a minority third (31.5%) preferred no strong robot personality factors. Humanoid robots also tended to be perceived as more intelligent than the mechanoid robots, but when combined with short height, were seen as less conscientious and more neurotic. The taller robots overall were also perceived as more human-like and conscientious than the short robots.

1 INTRODUCTION

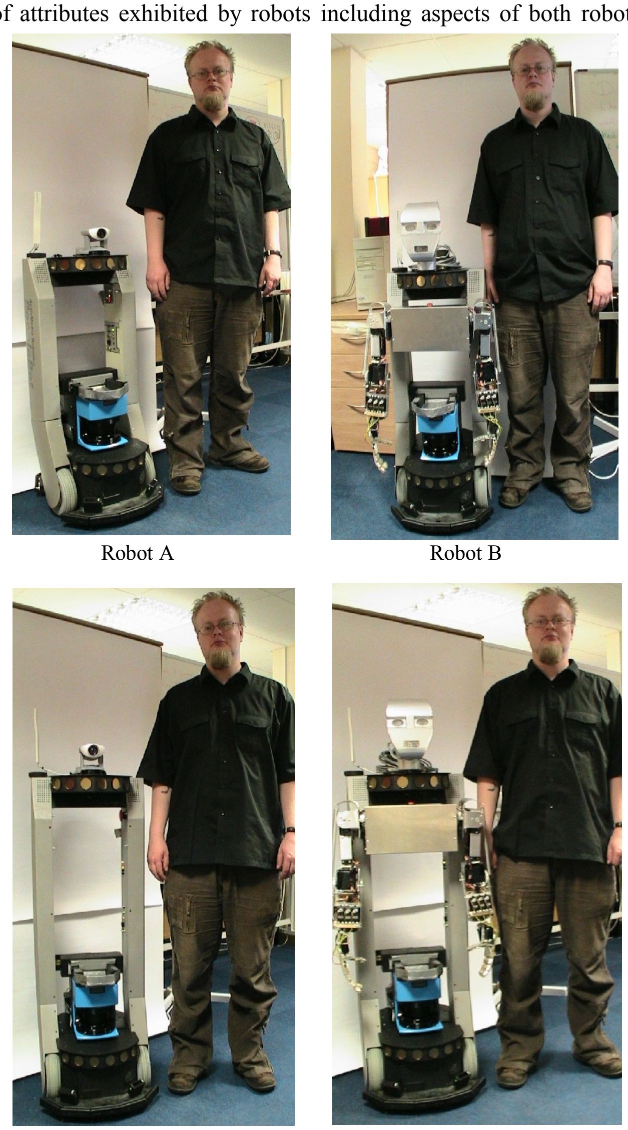
Within domestic environments, most current robots have mainly been seen as toys with (often limited) entertainment functions. These robots have usually exhibited a relatively small number of interaction functions and usually outwear their welcome after a relatively short time. In recent years the ongoing development of robot technical capabilities has enabled them to perform some useful functions such as simple cleaning tasks (eg. the ROOMBA vacuum cleaning robot), lawn mowing and basic (remote) security monitoring. However, these limited tasks have been selected for initial domestic robot applications specifically because they actually require little interaction with humans. Domestic robots in particular will exhibit a social aspect in most, if not all, interactions with humans. This is likely to be quantitatively and qualitatively different to that exhibited towards other technical artefacts [1] due to the physical embodiment of robots. We argue that if robots are to become truly useful in a human centered domestic environment they must satisfy two main criteria (cf. Dautenhahn et al. [2] & Syrdal et al. [3]):

1. *It must be able to perform a range of useful tasks or functions.*
2. *It must carry tasks or functions in a manner that is socially acceptable, comfortable and effective for people it shares the environment with and interacts with.*

1 The work described in this paper was conducted within the EU Integrated Projects COGNIRON ("The Cognitive Robot Companion") and LIREC (LIving with Robots and intEractive Companions) and was funded by the European Commission under Contract numbers FP6-002020 and FP7-215554.

2 All authors with the University of Hertfordshire, College Lane, Hatfield, Herts, AL10 5NG, United Kingdom. Email :{M.L.Walters, K.L.Koay, D.S.Syrdal, K.Dautenhahn, R.teBoekhorst,} @herts.ac.uk

Although many Human-Robot Interactions (HRIs) necessarily involve speech, our research emphasis is on the physical, spatial, visual and audible *non-verbal* social aspects of robots which must interact socially with humans. See Fong et al. [4] for an overview of robots designed to interact with humans in a social way. Peoples' social perceptions of robots may be affected by a number of attributes exhibited by robots including aspects of both robot



Robot A Robot B

Robot C Robot D

Figure 1: The PeopleBotTM Robots used for the HRI Studies: A) Short Mechanical, B) Short Humanoid, C) Tall Mechanical and D) Tall Humanoid. +

appearance and behaviour. With regard to robot appearances investigated here, we define Mechanoid and Humanoid robot appearances based on those for animated agents adopted by Gong & Nass [5] and of Android robots from MacDorman & Ishiguro [6]:

Mechanoid - relatively machine-like in appearance. In live and video based HRI trials described here, a robot described as mechanoid will have no overtly human-like features

Humanoid - a robot which is not realistically human-like in appearance and is readily perceived as a robot by human interactants. However, it will possess some human-like features, which are usually stylized, simplified or cartoon-like versions of the human equivalents, including some or all of the following: a head, facial features, eyes, ears, eyebrows, arms, hands, legs. It may have wheels for locomotion or use legs for walking.

Android - a robot which exhibits appearance (and behavior) which is as close to a real human appearance as technically possible.

Previously, both Khan [7] and Dautenhahn et al. [2] have found that people emphatically prefer domestic or service robots that are not realistically human-like in appearance. Therefore the robot appearances investigated here were limited to mechanoid and humanoid. Based on our previous experimental findings (cf. Woods et al. [8]) we speculated that the height of a robot may also affect peoples preferences. Therefore, peoples preferences and perceptions of robot height was also investigated in the current study.

1.1 BACKGROUND

Butler and Agar [9] explored the psychological effects of interactions between humans and mobile personal robots under conditions of different robot speeds, approach distances, and robot body design. Their experimental contexts included the robot approaching and avoiding a human, both while passing by and also performing non-interactive tasks in the same area as a human. Only direct, direct fast and indirect frontal approaches were considered. Two robot appearances were used; a tall humanoid robot (1.7m high with a simple head and arms) and a short mechanoid robot (0.35m high, cylindrical body), both with wheeled bases. Findings indicated that participants preferred closer (comfortable) approach distances by the short (0.35m) mechanoid robot than by the tall humanoid robot. Fast approaches (approx. 1m/s) by the tall humanoid robot in particular caused uncomfortable feelings in the human participants.

In previous human-robot comfortable approach distance experiments we have found most participants approached a mechanoid robot to distances that lie within the closer part of Hall's Personal Zone [10][11], reserved for conversation between friends (cf. Walters et al. [12],[13]). In another experiment (cf. Walters et al. [14]) we investigated comfortable human approach distances to a mechanoid robot which used four different voice styles. There were no significant differences found for comfortable approach distances for humans that had experienced a short previous interaction with a similar robot. However, non-habituated humans tended to approach a robot with a synthesized or female voice to further (comfortable) distances than to a robot with a male or no voice. A possible reason advanced for these initially greater approach distances were that they were due to slight initial uncertainty towards the robot, related to participants' initial expectations for robot appearance and robot voice style.

Our previous HRI proximity trials have also investigated robot to human approach distances for mechanoid (appearance) robots only. Others have investigated the effect of robot appearance on users' perceptions and expectations [15]. Minato et al. [16] and Goetz and Kiesler [17] have stressed the importance of consistency of robot appearance and behaviour with regard to forming and meeting humans expectations of appropriate social cues and technical capabilities. Lee & Kiesler [18] have found that people make very quick initial judgements of robots and their capabilities on very scant evidence or information, and particularly on robot appearance. Hinds et al. [19] also found that people treated machine-like robots in a more subordinate manner than more human-like robots. Walters et al. [20] found that people tended to rate particular robot behaviours or features less favourably when they are not perceived as consistent with the overall appearance of the robot

2 RESEARCH OBJECTIVES

The current study was performed as part of a larger series of HRI trials which took place in the University of Hertfordshire “Robot House”. The HRI trial series ran over five weeks with the main purpose of investigating how a group of long term participants' preferences and responses towards robots changed over that period. The main instrument to assess participants over the period was a controlled set of experiments which measured participants' ratings, responses and comfortable approach distance preferences towards a personal companion robot under a number of experimental conditions. The conditions which were controlled were robot appearance, robot height, task context, notions and perceptions of robot autonomy, and approach direction. A long term group had their responses and preferences tracked over five week period of habituation with the controlled set of trials repeated during the first, second and fifth week of the trial period. A greater number of short term participants underwent a controlled test series initially to establish a firm statistical baseline for comparison with the repeated test observations from the long term participants.

This paper presents outcomes specifically with regard to robot appearance and robot height preferences. Some aspects of the trials outcomes have been reported in Syrdal et al. [21] where findings indicate differences in approach direction preferences based on gender, that participants’ personality traits of extroversion and conscientiousness are associated with closer robot approach distance preference ratings, and differing perceptions and preferences for robot autonomy. Koay et al. [22] found that that preference ratings for robot approach direction and robot appearance changed over time. Participants who became accustomed to the robot tended to prefer to be more ‘in control’ of the situation - in that they appreciated reduced robot autonomy in case of unexpected events. The part of the trials, running between the second and fourth weeks, primarily to habituate the long term participants to the robots, also provided an opportunity to carry out a number of more exploratory experiments into different aspects of human and robot co-habituation. The results and data from this series of HRI mini-trials in weeks 2 to 4 of the trial series is currently being analysed and will be reported elsewhere. In order to investigate which of height, appearance or both factors influenced participants' preferences, ratings of robot behaviour and comfortable approach distances, a 2x2 combination of Tall/Short and Mechanoid/Humanoid robots were used in the trials (see Figure 1). Trial participants experienced interactions with just one out of the four possible robot appearance/height combinations (types). All participants completed post trial questionnaires where they were asked for their preferences and opinions with regard to all four possible robot appearances, height and their suitability for various tasks. For this study we advanced three hypothesis for testing:

1. *Participants' **preferences** for a tall or short robot will affect their robot to human proxemic distances*
2. *Participants' **preferences** for a mechanoid or humanoid robot appearance will affect their robot to human proxemic distances.*
3. *Participants will have a general overall preference for one (subjectively) optimal combination of robot appearance and height, based on their **perception** of robot personality factors and attributes.*

The responses to the post trial questionnaires are the main instruments of this study. The main aim was to investigate Mori's [23][24] observation that increasing the human-likeness of robots (but not to the extent that the "uncanny valley"repulsive effect was invoked) would improve users' interaction experience and effectiveness (cf. Goetz & Kiesler [17] and Minato et al. [16]). As none of the robots used in the study were particularly human-like in appearance, it was expected that the participants would generally prefer one of the more "humanoid" appearance robots. The robot height condition was incorporated in the HRI trials to investigate the notion that a shorter robot would be less intimidating and would therefore be allowed to approach closer than a taller robot. The findings for these trials reported previously in Syrdal et al. [21] indicated a general effect for mechanoid/humanoid robot appearance, whereby participants overall allowed a mechanoid appearance robot to approach more closely than the humanoid appearance robots. These findings also indicated that there were only significant differences in approach distance related to robot appearance, but not robot height. It was anticipated therefore, that robot appearance preferences may have effects on participants' preferred robot approach distances, but their *preferences* for robot height would have none.

3 EXPERIMENTAL METHOD

The HRI trials took place in a standard UK two bedroom rented apartment in order to provide a more ecologically valid setting. This "Robot house" has been used in our previous HRI trials (cf. Dautenhahn et al. [25], Woods et al. [8] and Koay et al. [26]). The territory of the trial is more neutral, home-like and realistic than a simulated environment in a laboratory or institutional setting. It was shown that this encourages participants to feel more at home and less scrutinized or judged, and thus more relaxed.

Twenty four Short Term participants carried out the controlled approach trials once only on a first exposure basis. The 12 Long Term participants carried out the controlled approach trials three times over the five weeks of the HRI trial series. The participants' ages ranged from 21 to 40. They were staff or students from various University of Hertfordshire departments, including Computer Science, Engineering, Psychology, Astronomy and Business Studies and not part of the HRI research team. The final questionnaire response data, which is the focus of this study was gained from participants only after all their live HRI trials were completed. In the case of the long term trial participants, this was after five weeks of exposure to the robot. As their responses will have been affected by the extended exposure, the data from the long term and short term groups of participants are considered separately and differences and comparisons are made between them where appropriate. The participants were drawn from the University population and were mainly postgraduate students, one academic staff member and one undergraduate student. Their ages ranged from 21 to 50 and there were 16 males, 9 females in the short term group, and 8 males, 4 females in the long term group. Participants were paid a modest compensation.

Four robot types were used for the HRI trials and differed only in the combination of the two controlled factors. The robots were carefully designed (using commercially available PeopleBots^{TM} robots as a common robot platform) to be the same in appearance and behaviour apart from the appearance and height factors.

Note, none of the robots used were particularly human-like in appearance. The terms "humanoid" and "mechanoid"are simply used here as labels as a shorthand to distinguish easily the main design features of the four robots (cf. Section 1).

Robot A was 1.2m tall and mechanical looking ("Short Mechanoid"), B was 1.2m tall and had a simple metallic head and two metallic human-like arms ("Short Humanoid"). C and D were both 1.4m tall, with C having mechanical features ("Tall Mechanoid") and D the same human-like features as B ("Tall Humanoid"). The terms "mechanoid" and "humanoid" were **not** used when talking to participants in the HRI trials or in questionnaires; The robots were simply referred to as Robots A, B, C or D (see Figure 1). All participants underwent the same controlled experiment with only one of the four robots types. The robot type actually used was assigned to each participant in sequence, so that approximately the same numbers experienced each robot type. (N=33; A, n=8; B, n=8; C, n=8; D, n=9), The participants used a Comfort Level Device (CLD, cf. Koay et al. [27]) to signal when the robot had approached to a distance which they found comfortable for each trial run, which was recorded from the robot's laser range sensor. The CLD was developed by the team especially to provide a means for participants to indicate uncomfortable situations by means of pressing a button on a wireless device which could be used to directly control the robot or log data as required. If a participant did not operate the CLD, the closest approach distance of the robot was recorded for the particular trial run.

To explore how the level of robot autonomy affected their comfortable approach distances, the CLD had two modes of operation which corresponded to the conditions Human in Control (HiC) and Robot in Control (RiC). Under the HiC condition, a press of the CLD button caused the robot to stop advancing towards the participants. Under the RiC condition, a press of the CLD button did not affect the robot's advance, and it carried on until the robot pre-programmed safety distance was triggered. In both cases the robot recorded the actual distance to the human, using the robot's internal laser range sensing system, when the CLD button was pressed. For each of the two robot autonomy conditions, three different task *context* conditions were studied: *No Interaction* - where the robot approached participants only incidentally while carrying out a task not involving the human. *Verbal Interaction* - where the robot approached participants in order to speak commands to the robot. *Physical Interaction* - where the robot approached the human for a joint task which required physical contact with the human. For each of the Interaction conditions, approaches were made from the front direct, and from the front right side quarter. These two approach directions (front and front side) were identified as most relevant in previous HRI trials (cf. Woods et al. [28] & [8]). Table 1 shows the experimental condition matrix of 2 (Autonomy) x 3 (Interaction Contexts) x 2 (approach Directions).

The main *relevant* findings of these HRI trials are briefly summarized here. Significant effects on comfortable approach distance were found for live robot appearance, but none for robot height. In general, people preferred the humanoid appearance robots (B and D) to keep a further distance away than the mechanical robots (A and C). Participants who rated highly on the Extroversion personality factor were associated with closer approach distance preferences than more introverted individuals, who preferred larger approach distances. In this previous analysis on the live HRI trial data, Syrdal et al. [29] found significant differences in comfortable approach distances for the Interaction context conditions, specifically between the Physical and Verbal Interactions, and the Physical and No Interaction contexts. For the purposes of the present study, a mean comfortable approach distances was aggregated for each participant over all their individual comfortable approach distances for all the experimental conditions. Post trial questionnaires were administered to participants and contained questions relating to participants' overall opinions, perceptions and preferences with regard to all four robot types from static photographs (Figure 1.). The four robot types (A, B, C or D) shown also included the one robot type which they had previously encountered in their live trials. The questions considered here were in two groups:

1. *Personal preference choices as to most and least liked robot types.*
2. *Subjective ratings of perceived attributes of the four robot types. This included ratings of robot personality factors (Big Five [30]), human-likeness and intelligence.*

The questions required the participants to respond in two possible ways.

Overall preference - a multiple choice selection response was presented (E.g. : "Which was your most preferred robot? Choose answer from: A, B , C or D:"). These nominal answers were used as grouping factors for a GLM (General Linear Model) Univariate ANOVA for significant differences between groups for mean comfortable approach distances (scale data).

Quantitative ordinal ratings - Used a five point Likert scale to obtain ratings (E.g. "How much did you like robot A?" Response from: 1 = Not at all, 2 = Not much, 3 = Neutral, 4 = A bit, 5 = A lot). These were compared with each other by non-parametric tests to obtain significant differences and correlations. Friedman ANOVA tests were used to test for significant differences between Likert [30] scale answers and Spearman Rho tests for significant correlations. Details of the particular questions relevant to this study are given in the appropriate part of the results section below:

Robot Autonomy	Interaction Context (P, V, and N) x Approach Direction (Front, Front Right)		
	Physical	**Verbal**	**None**
Robot in Control (RiC)	Front Front Right	Front Front Right	Front Front Right
Human in Control (HiC)	Front Front Right	Front Front Right	Front Front Right

Table 1: The controlled experimental conditions for comfortable approach distance studies

4 RESULTS

4.1 MOST PREFERRED ROBOT TYPE

Although most of the 24 short term (single exposure) participants preferred either the tall humanoid (Robot D, n=7, 36.8%) and short mechanoid robot types (Robot A, n=6, 31.6%), with smaller minorities preferring C (n =4, 21.1%) and B (n = 2, 10.5%). However, these numbers were not statistically significant (χ^2 = 3.105, df = 3, p = 0.376). The 12 long term participants, who had experienced five weeks of habituation before completing the questionnaires, showed similar proportions with robot D (n = 5, 42%) most preferred, with A and B (n = 3, 25%) joint second and C (n = 1, 8%), but again these were not statistically significant (χ^2 = 4.484, df = 3, p = 0.214). See Figure 2 for details.

Chi-square tests on the participants' preferred robot height (χ^2 = 0.290, df =1, p = 0.590) and appearance (χ^2 = 0.290, df =1, p=0.590) also indicated no overall significance. It must be therefore assumed that the reasons for a particular robot type (A, B, C or D) being preferred was based on participants' individual or internal preference factors.

Non-parametric tests also indicated there were no correlations (Spearman's Rho < 0.497, p > 0.190) between the robot types which short-term participants encountered in their HRI trials and their preferred robot types. No long term participant experienced actual interaction with their preferred robot type so similar tests could not be performed for this group. It seems therefore, that

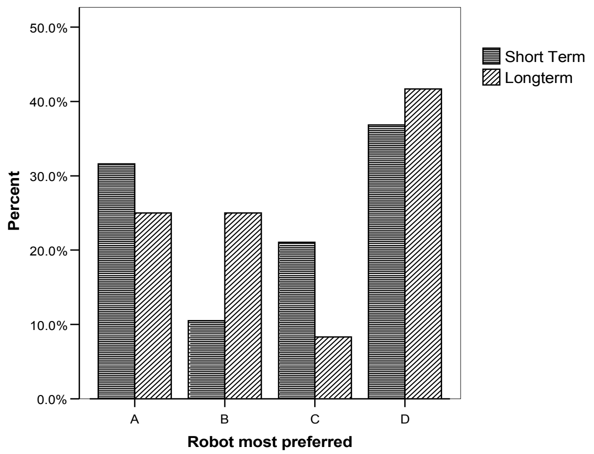


Figure 2: Robot types most preferred by first exposure (short term) and long term exposure participants. Robot A = Short Mechanoid, B = Short Humanoid, C = Tall mechanoid and D = Tall Humanoid

previous trial exposure to a particular robot type did not affect participants' preference for a particular robot type in any direct ways. Alternatively, participants might have chosen an appearance other than the one they encountered in live long-term studies since they might have hoped for a better performance of a 'new robot companion'. Given the current data, this explanation cannot be ruled out and requires further investigation. Note, ideally all long-term participants would have encountered twice a week each of the 4 robot types, but this was not possible for logistical reasons (i.e. entailing 96 HRI trials per week).

4.2 ROBOT TYPE PREFERENCES AND COMFORTABLE APPROACH DISTANCES

The four robots used for the live HRI trial runs, and in the still images shown to participants for the final questionnaires, were identical apart from the two factors of appearance and height. These factors were used as grouping factors for GLM (General Linear Model) Univariate ANOVA tests which examined the effects of participants' preferences on their comfortable approach distances from the live HRI trials with an actual robot. Syrdal et al. [29] found previously that the participants overall allowed the mechanoid robots to approach more closely than the humanoid robots. It was hypothesized that participants' preferences for robot appearance and height would also have an effect on comfortable robot approach distances overall (by any robot). The short term participants were considered separately from the long term sample, as the longer exposure of the long term sample to a particular robot may have caused systematic differences between the two sample sets. Table 2 summarizes the results obtained. The equivalent results for the long term participants are not directly comparable due to the smaller participant base and the repeated exposures to the robot over the five week period, but are given in Table 3. It can be seen that, as reported in Syrdal et al. [29] previously, the (live HRI trial) interacting robot's appearance has a significant effect on participants' approach distance preferences. When live robots are encountered, overall the participants preferred the humanoid robots to remain at further approach distances (mean = 0.645m, SD = 20.43) than the mechanoid robots (mean = 0.490m, SD = 20.43). The height of the live interacting robot had no significant effect.

However, for the short-term participants, there are significant effects related to their stated *preferences* for robots, both for preferred robot appearance ($p = 0.044$) and preferred robot height ($p = 0.003$), with the live interacting robot approach distances. There is also a significant interaction effect between most preferred robot appearance and actual robot appearance ($p = 0.002$). Participants who expressed a preference for humanoid and/or short robots, generally tolerated closer approaches by whichever robot they actually interacted with in the live HRI trials. In fact, the effect of the *preferred* height factor (variance = 31.5% of total) of the robot is slightly greater than that for the actual appearance of the interacting robot (variance = 25.5% of total). See Table 3 for a summary of these results.

The same UNANOVA with respect to the long-term participants must be treated with more caution (Table 3) due to the smaller sample set and the very different HRI trial and habituation procedures experienced. However, some broadly similar general trends can be tentatively identified. Although actual robot appearance is the only factor which is actually significant ($p = 0.038$), it can be seen proxemic effects related to preferred robot appearance and preferred robot height are both approaching significance at $p < 0.1$, and due to the small sample size (12) cannot be discounted as a possible real effect.

4.3 PARTICIPANT RATINGS OF ROBOT ATTRIBUTES AND PERSONALITY

The post trial questionnaires also asked participants to rate attributes of each of the robot types using five point Likert scales. Each robot was rated for degree of liking or disliking (Figure 3), and for attributes based on personality factors from the big five personality model, used commonly to rate human (cf. Goldberg

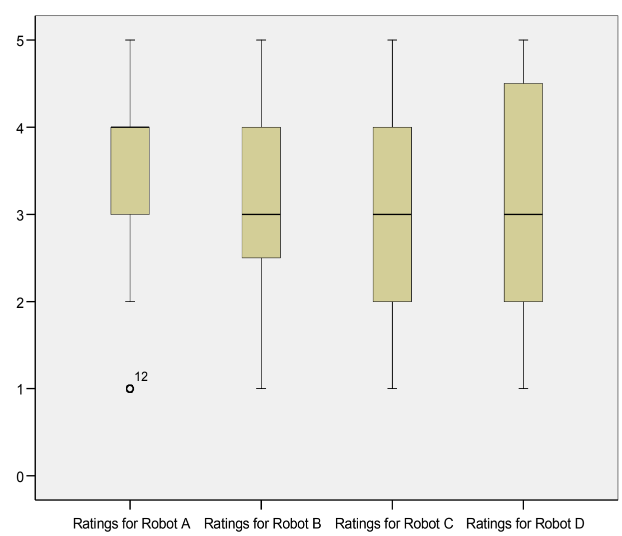


Figure 3: Mean participants' ratings of robot types (A, B, C, and D) on a five point Likert Scale (1 = Like not at all , 5 = Like very much).

[31]) and robot personality (cf. Syrdal et al. [21]). These factors were Extroversion, Agreeableness, Conscientiousness, Neurotism and Intelligence. An additional rating for perceived human-likeness of each robot type was included. The ratings for personality factors and human-likeness were tested using Spearman's Rho non-parametric tests for significant correlations with the overall participants' liking ratings for each robot type. These findings are discussed in detail below:

4.3.1 Overall Ratings of Robot.

Spearman's Rho tests found significant positive correlations between liking robot A and liking robot C (r = .532, p=.002), also between liking robot B and liking D (r = .609, p<.001). There was also a negative correlation between a liking for robot A and a disliking of robot D (r = -.392, p = .018). The common factor to these correlations is robot appearance. Individuals tend to like both the robots (B and D) with humanoid appearance, or like both the robots with mechanoid appearance (A and C). There were no significant correlations between the robot overall ratings and robot height, indicating that robot height did not have a major effect on participants' preferences for a particular robot type.

Factor	**Mean Square**	**Variance (%)**	**df**	**F**	**Sig.**
Most Preferred Robot Appearance	18365	8.5%	1	5.720	**0.044**
Most Preferred Robot Height	67832	31.5%	1	21.129	**0.002**
Live HRI Robot Appearance	54833	25.5%	1	17.080	**0.003**
Live HRI Robot Height	3799	1.7%	1	1.183	0.308
Most Preferred Robot Appearance + Live HRI Robot Appearance	69602	32.3%	1	21.680	**0.002**
Total Variance	215208	(99.5%)	11		

Table 2: GLM UANOVA test results for between subjects effects of *Short-Term* participants' preferences for robot appearance and height factors on comfortable robot approach distances.

Factor	**Mean Square**	**Variance (%)**	**df**	**F**	**Sig.**
Most Preferred Robot Appearance	8564	20%	1	15.418	0.059
Most Preferred Robot Height	6896	16.1%	1	12.415	0.072
Live HRI Robot Appearance	13713	32%	1	24.689	**0.038**
Live HRI Robot Height	1638	3.8%	1	2.950	0.228
Total Variance	42867	(79.1%)	11		

Table 3: GLM UANOVA test results for between subjects effects of *Long-Term* participants' preferences for robot appearance and height factors on comfortable robot approach distances.

4.3.2 Participants' Perceptions of Robot Types.

Robot A: Participants who liked short mechanoid robot A rated it as relaxed and contented (low neurotism) (r = -.445, p = .014) and also preferred mechanoid robot appearance (r = .517, p = .043).

Robot B: Participants who liked short humanoid robot B, preferred a humanoid robot appearance, perceived both humanoid robots B (r = -.420, p = .021) and D (r = -.517, p = .003,) as more extrovert, and perceived mechanoid robot A (r = .445, p = . 001) as less extrovert. They tended to rate both humanoid robots B (r = -.552, p = .002) and D (r = -.364, p = .048) as more agreeable, and tall mechanoid robot C (r = .508, p = .004) as less agreeable. They also rated short mechanoid robot B as being more intelligent and (surprisingly!) rated tall mechanoid robot C as more human-like (r = -.382, p = .037).

Robot C: Participants who liked mechanoid robot C, tended to perceive both mechanoid robots A (r = -.390, p = .033) and C (r = -606, p < .001) as more extrovert, robot C as more agreeable (r = -.398, p = .029) and low in neurotism (r = -.443, p = .014).

Robot D: Participants who liked tall humanoid robot D, especially preferred humanoid robots overall (r = .678, p < .001), saw robot D as more extrovert (r = -.605, p < .001), agreeable (r = -.393, p = .032), conscientious (r = -.433, p = .017), intelligent (r = -.513, p = .004) and human-like (r = -.449, p = .013). They also perceived mechanoid robot C as less agreeable (r = .589, p = .001) and humanoid robot B as more intelligent (r = -.430, p = .018), but did not rate B significantly for extroversion, agreeableness and conscientiousness.

Robot Height Overall: There were significant correlations for robot height preferences for the humanoid robots B (r =.367, p = .046) and D (r = -378, p = .040) which suggested that the taller humanoid robot was seen as conscientious, whereas the shorter humanoid robot B was rated as less conscientious.

Summary: It seems that most participants (68.5%) tended to perceive their preferred robot (B, C, and D) as having both extrovert and agreeable personalities. The minority of participants who preferred robot A (31.5%) seem to have perceived it as particularly lacking in any strong personality factors, apart from being relaxed and content (low neurotism personality factor rating). Participants who tended to prefer the more humanoid appearance robots, B and D (56.8%), seem to appreciate the generally stronger personality and intelligence factors which they are perceived as exhibiting. Robot D in particular also seems to be perceived as being more human-like, more conscientious and less stressful (low in neuroticism) than the shorter humanoid robot B.

5 DISCUSSION AND CONCLUSIONS

We have shown that peoples *preferences* for robot appearance and height are powerful indicators as to their likely responses to actual robots overall. This has implications for designers of domestic robots, as it implies that people who choose taller or humanoid type robots for their own domestic use, will tend to prefer closer approaches than those who choose a smaller mechanoid type robot design. These findings also suggest that a smaller more mechanoid robot appearance, with correspondingly less close approaches may be more acceptable for robots which must act within a public area. Although less obtrusive robots may be not be actively preferred by a majority of people, they may however be perceived as potentially less annoying or unsettling, and thus more acceptable by most people.

Although the tall humanoid robot was preferred overall, there was no significant overall preference for any particular one of the four robot types. Also, there was no overall preference for either of the factors for appearance (humanoid or mechanoid) or height (short or tall). There was some general tendencies for participants to prefer either the tall humanoid robot or the short mechanoid robot for both short and long term participants. It seems that individual participants had definite personal preferences, mainly based on their individual perceptions of the robot types however. These individual perceptions can be categorized into two main groups, a majority of roughly two thirds (68.5%) who preferred a robot which they personally perceived as having an extrovert and agreeable personality (primarily robot types B, C and D), and a minority third (31.5%) who actually preferred a small robot with no strongly perceived robot personality factors (robot type A). One might consider that robot type A may have been preferred by those participants who just wanted an unobtrusive, emotionally undemanding servant or smart machine. Syrdal et al. [21] found a possible relationship between peoples personalities and their preferences. More introverted individuals tended to prefer mechanoid robot appearance and extroverts prefered more humanoid robots. This suggests that peoples' preferences for robot appearance and behaviour may also be related to aspects of their personalities. However, more specific focussed research would be needed to confirm and investigate this aspect further.

The humanoid robots (B and D) also tended to be perceived as more intelligent with richer personalities than the mechanoid robots. However, When humanoid appearance was combined with short height (B), the robot also tended to be perceived as less conscientious and more neurotic. Interestingly, these are traits that are typical of human children and therefore possibly it may have been perceived as requiring more close attention and supervision, and be less responsible. The tall robots (even mechanoid robot C) overall also tended to be perceived overall as more human-like and conscientious than the short robots. It may be that the participants perceived the small stature of the humanoid robot as more childlike, while seeing the taller humanoid as more adult. Therefore the overall popularity of the tall humanoid robot D may be due to the participants' perceptions that it looked like it could actually carry out tasks responsibly and cheerfully. The low overall preference rating of short humanoid robot B may be due to participants perception that it may be childlike and would therefore not be capable of carrying out useful work effectively.

The tall mechanoid robot C was rated by some participants as human-like, and it may have been most preferred by some participants as they personally perceived it as exhibiting extrovert and agreeable personality factors.

The suggested explanations we have provided above regarding people's choices and preferences need to be investigated further in future work. These results have provided some insights into how humans perceive and rate robots on initial acquaintance. More data from this study still awaits analysis. How participants own personality factors influences their preferences and perceptions of the four robots, and analysis of the participants views on task domains, capabilities and suitabilities of the four robots are left for presentation in future papers.

6 REFERENCES

[1] C. Nass, J. Staeur & E. Tauber, 'Computers are Social Actors', *Proceedings of Conference of Human Factors in Computing Systems, Human-Computer Interaction (CHI 94)*, Boston, Massachusetts, USA, 72-78, (1994).

[2] K. Dautenhahn, S. N. Woods, C. Kaouri, M. L. Walters, K. L. Koay & I. Werry, 'What is a Robot companion - Friend, Assistant or Butler?', *Proceedings of IEEE RSJ International Conference on Intelligent Robot Systems (IROS'05)*, Edmonton, Canada, 1488-1493, (2005).

[3] D. S. Syrdal, K. Dautenhahn, S. N. Woods, M. L. Walters, K. L. Koay, 'Doing the Right Thing Wrong' - Personality and Tolerance to Uncomfortable Robot Approaches', *Proceedings of The 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06)*, University of Hertfordshire, UK, 183-188, (2006).

[4] T. Fong, I. Nourbakhsh, K. Dautenhahn, 'A Survey of Socially Interactive Robots', *Robotics and Autonomous Systems*, *42(4-3)*, 143–166, (2003).

[5] L. Gong, C. N. Nass, 'When a Talking-Face Computer Agent is Half-Human and Half-Humanoid: Human Identity and Consistency Preference', *Human Communication Research*, *33(2)*, 163-193, (2007).

[6] K. MacDorman, H. Ishiguro, 'The Uncanny Advantage of Using Androids in Coognitive and Social Science Research ', *Interaction Studies*, *7(3)*, 297-337, (2006).

[7] Z. Khan, '*Attitude towards intelligent service robots*', , TRITA-NA-P9821, , NADA, KTH (1998).

[8] S. N. Woods, M. L. Walters, K. L. K.oay, K. Dautenhahn, 'Methodological Issues in HRI: A Comparison of Live and Video-Based Methods in Robot to Human Approach Direction Trials', *Proceedings of The 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06)*, Hertfordshire, UK, 51-58, (2006).

[9] J. T. Butler, A. Agah, 'Psychological Effects of Behavior Patterns of a Mobile Personal Robot', *Autonomous Robots*, *10*, 185-202, (2001).

[10] E. T. Hall, *The Hidden Dimension*, Doubleday, NY, (1966).

[11] E. T. Hall, 'Proxemics', *Current Anthropology*, *9(2-3)*, 83-108, (1968).

[12] M. L. Walters, K. Dautenhahn, K. L. Koay, C. Kaouri, R. te Boekhorst, C. L. Nehaniv, I. Werry, D. Lee, 'Close Encounters: Spatial Distances Between People and a Robot of Mechanistic Appearance', *Proceedings of IEEE-RAS International Conference on Humanoid Robots (Humanoids2005)*, Tsukuba, Japan, 450-455, (2005).

[13] M. L. Walters, K. L. Koay, S. N. Woods, D. S. Syrdal, K. Dautenhahn, 'Robot to Human Approaches: Comfortable Distances and Preferences', *Proceedings of the AAAI Spring Symposium on Multidisciplinary Collaboration for Socially Assistive Robotics, (AAAI SS07-2007)*, Stanford University, Palo Alto, CA, USA, , (2007).

[14] M. L. Walters, K. L. Koay, K. Dautenhahn, R. te Boekhorst & D. S. Syrdal, 'Human Approach Distances to a Mechanical-Looking Robot with Different Robot Voice Styles', *Proceedings of the 17th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2008)*, Munich, Germany, 707-712, (2008).

[15] T. Kanda, T. Miyashita, T. Osada, Y. Haikawa, H. Ishiguro, 'Analysis of Humanoid Appearances in Human Robot Interaction', *Proceedings of IEEE/RSJ Intelligent Robot Systems (IROS 2005)*, Edmonton, Canada, 899-906, (2005).

[16] T. Minato, K. F. MacDorman, M. Shimada, S. Itakura, K. Lee, H. Ishiguro, 'Evaluating Humanlikeness by Comparing Responses Elicited by an Android and a Person', *Proceedings of the Second International Workshop on Man-Machine Symbiotic Systems.*, Kyoto, Japan., 373-383, (2004).

[17] J. Goetz, S. Kiesler, 'Cooperation with a Robotic Assistant', *Proceedings of the Conference on Human Factors in Computing Systems (CHI'02)*, New York, USA., 578 - 579, (2002).

[18] S. Lee, I. Yee-man Lau, S. Kiesler & C. Chiu, 'Human Models of Humanoid Robots', *Proceedings of the 2005 International Conference on Robotics and Automation (ICRA 05)*, Barcelona, Spain, 2767- 2772, (2005).

[19] P. J. Hinds, T. L. Roberts, H. Jones, 'Whose Job Is It Anyway? A Study of Human-Robot Interaction in a Collaborative Task', *Human Computer Interaction*, *19*, 151-181, (2004).

[20] M. L. Walters, D. S. Syrdal, K. Dautenhahn, R. te Boekhorst, K. L. Koay, 'Avoiding the Uncanny Valley – Robot Appearance, Personality and Consistency of Behavior in an Attention-Seeking Home Scenario for a Robot Companion', *Journal of Autonomous Robots*, *24(2)*, 159-178, (2008).

[21] D. S. Syrdal, M. L. Walters, K. L. Koay, S. N. Woods & K. Dautenhahn, 'Looking Good? Appearance Preferences and Robot Personality Inferences at Zero Acquaintance', *AAAI - Spring Symposium 2007, Multidisciplinary Collaboration for Socially Assistive Robotics*, Stanford University, Palo Alto, California, 86-92, (2007).

[22] Koay, K. L., Syrdal, D. S., M. L. . Walters, K. Dautenhahn, 'Living with Robots: Investigating the Habituation Effect in Participants' Preferences During a Longitudinal Human-Robot Interaction Study', *Proceedings of the 16th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2007)*, South Korea, 564-569, (2007).

[23] M. Mori, 'Bukimi No Tani (The Uncanny Valley)', *Energy*, *7(4)*, 33-35, (1970).

[24] K. F. MacDorman, 'Androids as an experimental apparatus: Why is there an uncanny valley and can we exploit it?', *Proceedings Of the CogSci 2005 Workshop: Toward Social Mechanisms of Android Science*, Stresa, Italy, 106-118, (2005).

[25] K. Dautenhahn, M. L. Walters, S. N. Woods, K. L. Koay, C. L. Nehaniv, E. A. Sisbot, R. Alami, T. Simeon, 'How may I serve you? A robot companion approaching a seated person in a helping context', *Proceedings of ACM SIGCHI/SIGART 2nd Conference on Human Robot Interaction (HRI '06)*, Salt Lake City, Utah, USA, 172-179, (2006).

[26] K. L. Koay, E. A. Sisbot, D. S. Syrdal, M. L. Walters, K. Dautenhahn, Alami R., 'Exploratory Studies of a Robot Approaching a Person in the Context of Handing Over an Object', *Proceedings of AAAI - Spring Symposium 2007: SS07, Multidisciplinary Collaboration for Socially Assistive Robotics , AAAI Technical Report*, Stanford University, Palo Alto, Ca, USA, 18-24, AAAI Press (2007).

[27] K. L. Koay, M. L. Walters, S. N. Woods, Dautenhahn K., 'Empirical Results from Using a Comfort Level Device in Human-Robot Interaction Studies', *Proceedings of ACM SIGCHI/SIGART 2nd Conference on Human Robot Interaction (HRI '06)*, Salt Lake City, Utah, USA, 194-201, (2006).

[28] S. N. Woods, M. L. Walters, K. L. Koay, K. Dautenhahn, 'Comparing Human Robot Interaction Scenarios Using Live and Video Based Methods: Towards a Novel Methodological Approach', *Proceedings of The 9th International Workshop on Advanced Motion Control (AMC'06)*, Istanbul, Turkey, 750-755, (2006).

[29] D. S. Syrdal, K. L. Koay, M. L. Walters, K. Dautenhahn, 'A Personalised Robot Companion? - The Role of Individual Differences on Spatial Preferences in HRI Scenarios', *Proceedings of the 16th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2007)*, Korea, 26-29, (2007).

[30] R. Likert, 'A Technique for the Measurement of Attitudes', *Archives of Psychology*, *140*, , (1932).

[31] L. R. Goldberg, 'The Structure of Phenotypic Personality Traits', *American Psychologist* , *48*, 26-34, (1993).

An Empirical Framework for Human-Robot Proxemics.[1]

Michael L Walters, Kerstin Dautenhahn, René te Boekhorst, Kheng Lee Koay, Dag Sverre Syrdal and Chrystopher L. Nehaniv.[2]

Abstract. An empirical framework for Human-Robot (HR) proxemics is proposed which shows how the measurement and control of interpersonal distances between a human and a robot can be potentially used by the robot to interpret, predict and manipulate proxemic behaviour for Human-Robot Interactions (HRIs). The proxemic framework provides for incorporation of inter-factor effects, and can be extended to incorporate new factors, updated values and results. The framework is critically discussed and future work proposed.

1 INTRODUCTION AND BACKGROUND

If domestic and service robots are to become truly useful, in addition to performing useful tasks they must also be socially acceptable and effective when interacting with the people who share their working environment (cf. [1], [2] & [3]). Fong et al. [4] describe socially interactive robot characteristics: To express and/or perceive emotions, communicate with high-level dialogue, learn and/or recognize models of other agents, establish and maintain social relationships, use natural cues (gaze, gestures, etc.), exhibit distinctive personality and character, and learn or develop social competencies.

Nass et al. [5] and Reeves and Nass [6] found that people have social relationships with computers including politeness, reciprocity, attribution of gender stereotypes and personality in spite of knowing that they are machines. Therefore, people will react and relate socially to robots in some of the ways that they do to humans, computers and other artefacts. Embodied non-verbal interactions, such as approach, touch, and avoidance behaviours, are fundamental to regulating human-human social interactions (cf. [7]) and the physical embodiment of robots makes it likely that they will have to exhibit appropriate non-verbal interactive behaviours. The study of how humans use and manipulate distances between each other with regard to social behaviour and perceptions is called *proxemics,* This paper focuses on empirical research into Human-Robot (HR) proxemics and proposes a framework for HR proxemic factors which will facilitate future study. Relevant findings from Human-Human proxemics, Human-Computer Interaction (HCI) and Human-Robot Interaction (HRI) research are first reviewed.

[1]The work described in this paper was conducted within the EU Integrated Projects COGNIRON ("The Cognitive Robot Companion") and LIREC (LIving with Robots and intEractive Companions) and was funded by the European Commission under contract numbers FP6-002020 and FP7-215554.

[2]All authors with the University of Hertfordshire, College Lane, Hatfield, Herts, AL10 5NG, UK. Email: {M.L.Walters, K.Dautenhahn, R.teBoekhorst, K.L.Koay, D.S.Syrdal, C.L.Nehaniv}@herts.ac.uk

1.1 Proxemics

Harrigan et al. [8] provide a general introduction into non-verbal human behaviour, including proxemics. Hall [10] observed that human social spatial distance varies by the degree of familiarity between interacting humans and the number of interactors. Later, Hall [11] provided a framework which categorized the main social spatial zones by interaction and situation. Hall estimated these distances visually in terms of arm lengths, close contact and threat/flight distances, but other researchers have assigned more precise numerical values [9] (Table 2). Other factors have been identified which affect human-human proxemics and there is evidence that even relatively small differences and changes in human proxemic distances (of the order of 2 cm to 15 cm) are significant. Horowitz et al. [12] found that participants were "comfortable" approaching arbitrarily close to inanimate objects. Sommer [13] stated that there were no social proxemic effects for objects and people can approach arbitrarily close without discomfort. Stratton et al. [14] found that mean "comfortable" approach distances between human participants were approximately 51cm (20in) and that significant differences in approach distance correlated with participants' notions of "self-concept" which is a trait that is related to the social status of participants. They also found that participants approached a dressed headless tailor's dummy to a mean approach distance of 55cm (22in). This was slightly (but not significantly) greater than the mean human-human approach distances. The authors suggested that participants may have taken a slightly greater approach distance due to a mild form of the "fear of the strange" effect. This was observed in animals and noted by Hebb [15], where chimpanzees were observed to keep greater distances from images of distorted chimpanzee faces and limbs than they did to non-distorted parts or other images. Kubinyi et al. [16] found a similar effect with regard to dogs and robots; adult dogs tended to leave larger distances between themselves and a "furry robot" dog than for either a toy car (the control), a real puppy or a "hard robot" dog. This effect is possibly related to the biological

Range	Situation	Personal Space Zone
0 to 0.15m	Lover or close friend touching	Intimate Zone
0.15m to 0.45m	Lover or close friend only	Close Intimate Zone
0.45m to 1.2m	Conversation between friends	Personal Zone
1.2m to 3.6m	Conversation to non-friends	Social Zone
3.6m +	Public speech making	Public Zone

Table 1 Human-Human Personal Space Zones (cf. Lambert [9])

origins of the "uncanny valley" effect in humans noted by Mori [17] and discussed by Brenton et al. [18] and MacDorman [19] with regard to androids (very human-like) robots.

Perceived threat can also affect proxemic distances and is possibly related to the "flight reaction" originally observed in birds by Hediger [20]. This occurs when a perceived threat rises beyond a certain level and an animal will prepare to either fight or flee according to its nature and the context of the threat. For humans and primates where the perceived threat is actually minimal (i.e. feeling uncertain rather than threatened), the response is proportional and they take up slightly greater distances from the source of the perceived potential "threat".

Gillespie and Leffler [7] concluded that much of the observed variation in social distance between *communicating* humans is accounted for by the relative status of the interactants. The higher the relative status of one interactor, the more distance relatively low status interactors will keep, whereas relatively high status individuals do not respect the social spaces of other lower status individuals. The concept of status is not a one dimensional quantity: It can be perceived in terms of a combination of factors including age, hierarchical seniority, self-concept, intelligence, charisma, physical presence, gender and force of personality.

Burgoon and Jones [21] explained many seemingly contradictory aspects of human-human proxemic behaviour by suggesting that relatively small manipulations of the distance between interactants were a social "reward and punishment" mechanism. In any interaction there would be an optimal social distance and that one or other of the interactors could then "punish" or "reward" the other interactor by making (relatively small) adjustments in an appropriate direction. For example, if a woman wanted to encourage a man's attention she may "reward" him by moving closer than might be expected or, on the other hand, literally "keep her distance". The same theory can also explain how high status interactors can "reward" lower status interactors by moving closer, but lower status interactors can "reward" higher status interactors by keeping a greater distance.

In the related field of HCI Benford et al. [22] used a spatial zone model to detect the willingness of avatars to interact with agents. Bailenson et al. [23] investigated interpersonal distances in immersive virtual environments between humans (avatars) and computers (agents). Overall, participants maintained interpersonal distances that were comparable to those for real humans (approximately 0.5m), keeping greater distances from virtual humans when approaching their fronts compared to their backs. When participants *believed* that virtual humans were avatars (even when really computer-controlled) they also tended to keep further interpersonal distances than when they believed the agents were computer controlled. The significant differences between interpersonal distances found by this study were of the order of 3 to 7 cm (approx. 1.4 to 2.8 inches).

1.2 Human-Robot Proxemics

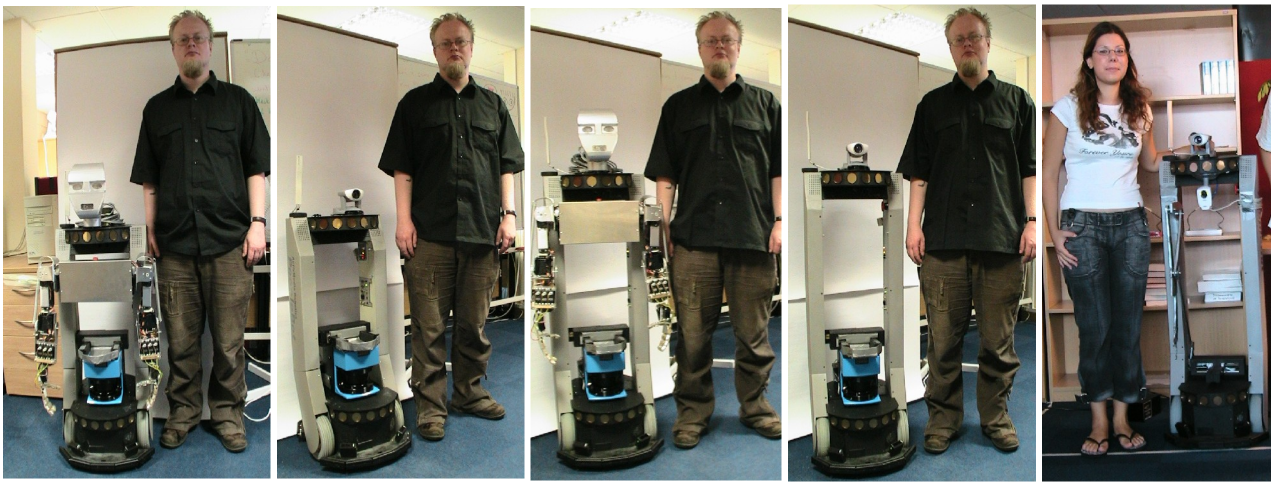
Breazeal [24] found that humans responded socially to expressive zoomorphic robots in some very fundamental non-verbal ways with regard to turn-taking in speech communication and respecting the robot's interpersonal space. Nomura et al. [25] found that both participants' negative attitudes and anxiety towards a small size humanoid robot, RobovieM (29 cm tall and 1.9 kg), had statistically significant effects on users preferred (comfortable) robot approach distances. Hűttenrauch et al. [26] concluded that in HRI user trials most participants kept interpersonal distances from the robot corresponding to Hall's Personal Spatial Zone (0.45m to 1.2m). In initial HRI proxemic trials we found that groups of children tended to approach a PeopleBot™ robot to similar distances on first encounter [27] but for individual adults approaching the same robot, the approach distances were more ambivalent and inconclusive [28][29]. In an HRI experiment with a similar mechanoid appearance robot with a simple pointing arm (Figure 1, Robot E) which used different voice styles, participants initially encountering the robot took significantly different comfortable approach distances [30]. It was suggested that these differences may be caused by a slight initial uncertainty due to the perceived inconsistency between the robot appearance and voice styles (Table 2).

Our recent main series of HRI trials run in 2006, have found factors affecting Robot to Human (RH) comfortable approach distances which are summarized here (Table 2). Participants generally allowed robots to approach more closely during physical interactions than under verbal or no interaction conditions [31]. People generally preferred more humanoid appearance robots to keep a further distance away than mechanoid robots, but the robot height (short = 1.2m, tall = 1.4m) had no effect [3]. Interestingly however, we found that participants' *preferences* for particular robot attributes did affect participants' comfortable approach distances with regard to all robot types [32]. Those who *preferred* a humanoid robot (i.e. with some human-like features, but obviously robotic) appearance (Figure 1, Robots B and D) also tended to allow whichever live robot they were interacting with to approach closer than those who *preferred* a mechanoid robot. Also, those who *preferred* a tall robot (Figure 1, Robots C and D), also tended to allow whichever live robot type they were interacting with to approach closer than those who *preferred* the short robot.

The results from our trials are summarized in Table 2 where all distances have been compensated to satisfy a standard measurement between the human and the robot's closest body trunk parts (i.e. not including arms or manipulators). These distance measurements (as best as we can tell from the published details) are also roughly comparable to those by Hall for his spatial zone distances and Stratton et al. [14]. These HRI trials were run using semi-autonomous robot control techniques in resource intensive HRI trials [33].

2 A PROVISIONAL FRAMEWORK FOR HR PROXEMICS

Table 3 shows the factors for robot appearance, preferences and interaction context and situation, which we have found experimentally to affect HR approach distances. The distance measurements are rounded to the nearest whole cm. Some values are predicted estimates from based on our earlier experimental results which indicated a relatively high degree of symmetry between similar physical situations where Humans approached Robots and Robots approached Humans for comfortable approach distances [29]. The distances are given as relative differences to the Grand Mean approach distance of 57cm. This



Robot A *Robot B* *Robot C* *Robot D* *Robot E*

Figure 1 The PeopleBot™ Robots used for the large HRI Studies: A) Short Mechanoid, B) Short Humanoid, C) Tall Mechanoid, D) Tall Humanoid. and E) the Mechanoid robot used for the robot voice style trial.

figure was gained from a previous large scale HRI trial, and was calculated over all the repeated measured preferred approach distances for all the trial conditions (for robot autonomy, interaction context, situation and approach direction). It is also close to the overall mean approach distances obtained by Stratton et al. for both humans (20in = 51cm) and the tailors dummy (22in = 56cm) used as a control [14].

Using the relative differences given in Table 3, a *default* approach distance estimate can be calculated for a robot encountering any combination of proxemic factors in the first column. For example, consider the case where a Humanoid robot approaches a human to hand over an object. Note the factors which apply, then calculate the default approach distance for the particular situation and context. In this case, the distance would be: (Base distance =) 57cm + (Humanoid-RH Approach =) 3cm - (Giving Object RH Approach=) 7cm = 53cm. If other any other factors are known (e.g. if the preferred height was short, then adjust by -1cm), then they can also be incorporated into the calculation. As other factors which affect HR proxemics become known or quantified, they can be incorporated into the model and used to refine or extend the applicability of the proxemic estimates produced. For example, the robot's voice style has already been shown to affect HR proxemics [30] and it is likely that gender and gestures by both human and robot may well affect HR proxemic behaviour, as is the case for human proxemics [8].

Approach Context	Mean (cm)	Standard Error (mm)	95% Confidence Interval (cm)	
			Lower Bound	*Upper Bound*
Grand Mean	57	18.312	53.04	60.50
Interaction: Pass	60	13.055	57.60	62.73
Verbal	60	13.055	58.00	63.09
Physical	49	13.055	46.28	51.4.1
Appearance:				
Mechanoid	51	10.830	48.71	52.98
Humanoid	62	10.486	60.11	64.24
Control: Robot	57	18.870	53.39	61.07
Human	56	21.069	52.02	60.60
Direction: Front	58	20.510	54.12	62.47
Side	55	18.433	51.44	59.04
Preferences:				
Mechanoid	60	17.393	46.80	54.22
Humanoid	56	17.946	61.57	69.22
Short	61	18.349	53.56	61.3.8
Tall	55	16.967	54.81	62.0.5
Initial Uncertainty	71	67.770	57.04	84.27

From Koay et al. [31], Syrdal et al. [3], Walters et al. [30] and Walters [34].
Note: These values have been compensated to make the distance measures directly comparable.

Table 2 RH Approach Distances vs. Interaction Context

If a particular factor is not known, then it is wise to err on the side of caution and assume that the furthest distance would apply. An approach that was was too close might be interpreted as invading personal space, while an approach that was slightly too far away would be perceived as keeping a respectful distance. For example, if a person's preference for height is not known, it is safest for the robot to assume that their preference is for small robots as this would ensure that any error in approach distance positioning by the robot would result in an approach distance that would be further away than might actually be preferred. It should also be straightforward to incorporate (modified) rules, with appropriate weightings for Hall's social and public spatial zone distances to provide for appropriate proxemic behaviour by the robot over larger distances in open areas and for different physical situations [35]. The framework also lends itself to incorporating other different scales for the rating of robot appearance (e.g. realistic-iconic, realistic-abstract or machine-organic dimensions, cf. [36] & [37]).

This method assumes that the HR proxemic factors are linear

Factor	Situation(s)	Context(s)	Base Distance = 57cm Estimated Adjustment for Factor (± 0.5cm)
Attribute or Factor of Robot			
Mechanoid Robot	RH Approach HR Approach	All	-3 -7
Humanoid Robot	RH Approach HR Approach	All	+3 -1
Verbal Communication	RH Approach	Verbal Interaction	+3
Giving object	RH Approach	Physical Interaction	-7
Taking object	RH Approach	Physical Interaction	-7?
Passing	RH Approach	No Interaction	+4
Direction from:	RH Approach	Front Right/Left	+2 -2
Attribute or Factor of Human			
Preferred Robot Humanoid	RH Approach	All Private	-3
Preferred Robot Mechanoid	RH Approach	All	+3
Preferred Height Tall	RH Approach	All	-1
Preferred Height Short	RH Approach	All	+2
Uncertainty or perceived Inconsistency	HR Approach	Initial Encounter	+13
Verbal Communication	HR Approach	Verbal Interaction	+3
Giving object	HR Approach	Physical Interaction	-7?
Taking object	HR Approach	Physical Interaction	-7?
Passing	HR Approach	No Interaction	+4

? *Indicates an estimated value based on the observation from our earlier study [28][29] that RH and HR approach distances were highly correlated and exhibited a high degree of symmetry between HR and RH approaches.*

Table 3 Factors Affecting HR proxemics

and independent. However, the number of robot types studied here is too few to make any conclusions as to the form (linear or otherwise) of the relationships between the factors examined (e.g. robot appearance) and the precise numerical value of their effects. There are also indications that some of the factors are dependent on each other. For example, from [38] it was found that the preferred robot appearance and actual robot appearance factors have a combined effect on HR approach distances (Table 2). In this case a practical approach would be to apply a correction if both factors are present, possibly by means of a look up table. It should be noted that few real world systems actually exhibit linear behaviour, but often by assuming a linear response, a reasonably precise control output can be obtained without having to implement more sophisticated non-linear control methods.

The values provided in Table 3 are obtained from our controlled HRI trials and the numbers of participants for the relevant experiments are relatively large compared to those typically found in HRI trials of this type. However, it is desirable to perform approach distance experiments with a much larger experimental sample, and with a large number of robots with different appearance and behaviour attributes to properly establish the range, form and parameters of HR proxemics. To perform the required number of experimental runs, it will be necessary to implement fully autonomous robot control and automatic data collection methods for future experiments.

Another issue is the large variance observed for both the Grand Mean and for the marginal means in these samples (Table 2). This suggests that individual differences between participants play a large role in determining proxemic preferences in any given instance of an HRI encounter. This makes the study of systematic variations in proxemic preferences according to measurable individual differences factors, such as personality and demographic data, as well as more HRI specific factors like the NARS [39] or UTAUT [40] scales, a salient avenue of investigation in order to establish a HR proxemic framework with greater predictive power.

3 IMPLEMENTATION OF A HUMAN-ROBOT PROXEMIC SYSTEM

In order to test, verify and extend the application range of the empirical HR proxemic framework, the next stage would be to conduct live HRI experiments, with a HR proxemic control system based on the empirical framework implemented on a range of robot platforms. Mitsunaga et al. [41] implemented an adaptation mechanism based on policy gradient reinforcement learning for robot proxemic and gaze behaviours using initial default parameters for HR proxemics based on Hall's social spatial zones. Their system illustrates the viability of using an adaptive control system to refine initial default values for proxemics for particular HRIs based on empirically obtained default values. However, we propose that a prototype implementation using a fuzzy logic based control system might be particularly well suited for verification and further research purposes. The various HR proxemic factors could be incorporated by means of fuzzy rule sets. The weightings of the factors could then be dynamically "tuned" by means of a number of well known learning algorithms (cf. [42] [43]), possibly using actual real time user feedback (cf. [44]). This would provide a learning mechanism so the robot could effectively adapt its proxemic behaviour to individual users preferences and requirements. The advantage of a fuzzy logic based control system is that as the robot becomes acclimatised to the proxemic preferences of more users, contexts and situations, it is possible to interrogate the fuzzy system proxemic factor weightings, and thus work back to estimate and explore the relationships between HR proxemics and the factors. Fine adjustments of human-robot interpersonal distances according to a number of observed factors (as proposed by Walters et al. [28]) related to the internal qualities of the interacting humans, intrinsic robot attributes, the external physical situation and the task context, is a worthwhile contribution towards the goal of a robot companion that can be individualized, personalized and will adapt itself to the user as suggested by Dautenhahn [45].

ACKNOWLEDGEMENTS

Many thanks also to all our colleagues who have provided help and support in running the live HRI studies. In particular we would like to thank David Lee, Iain Werry, Christina Kouri and Sarah Woods for their contributions in the early stages of developing and running the various HRI trials which have provided data for this paper.

REFERENCES

[1] K. Dautenhahn, S. N. Woods, C. Kaouri, M. L. Walters, K. L. Koay & I. Werry, 'What is a Robot companion - Friend, Assistant or Butler?', *Proceedings of IEEE RSJ International Conference on Intelligent Robot Systems (IROS'05)*, Edmonton, Canada, 1488-1493, (2005).

[2] S. N. Woods, K. Dautenhahn, C. Kaouri, R. te Boekhorst, K. L. Koay, M. L. Walters, 'Are Robots Like People? - Relationships between Participant and Robot Personality Traits in Human-Robot Interaction', *Interaction Studies*, *8(3)*, 281-305, (2007).

[3] D. S. Syrdal, K. L. Koay, M. L. Walters, K. Dautenhahn, 'A Personalised Robot Companion? - The Role of Individual Differences on Spatial Preferences in HRI Scenarios', *Proceedings of the 16th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2007)*, Korea, 26-29, (2007).

[4] T. Fong, I. Nourbakhsh, K. Dautenhahn, 'A Survey of Socially Interactive Robots', *Robotics and Autonomous Systems*, *42(4-3)*, 143–166, (2003).

[5] C. Nass, J. Staeur & E. Tauber, 'Computers are Social Actors', *Proceedings of Conference of Human Factors in Computing Systems, Human-Computer Interaction (CHI 94)*, Boston, Massachusetts, USA, 72-78, (1994).

[6] B. Reeves, C. Nass, *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*, Cambridge University Press, (1996).

[7] D. L. Gillespie, A. Leffler, 'Theories of Non-Verbal Behavior: A Critical Review of Proxemics Research', *Socialogical Theory*, *1*, 120-154, (1983).

[8] J. A. Harrigan, K. R. Scherer, R. Rosenthal, *The New Handbook of Methods in Nonverbal Behavior Research*, Oxford University Press, (2005).

[9] D. Lambert, *Body Language.*, Harper Collins, London, (2004).

[10] E. T. Hall, *The Hidden Dimension*, Doubleday, NY, (1966).

[11] E. T. Hall, 'Proxemics', *Current Anthropology*, *9(2-3)*, 83-108, (1968).

[12] M. J. Horowitz, D. F. Duff, L. O. Stratton, 'Body Buffer Zone', *Archives of General Psychiatry*, *XI*, 651-656, (1964).

[13] R. Sommer, *Personal Space: The Behavioral Basis of Design*, Prentice Hall, New Jersey, (1969).

[14] L. O. Stratton, D. J. Tekippe, G. L. Flick, 'Personal Space and Self Concept', *Sociometry*, *36*, 424-429, (1973).

[15] D. O. Hebb, *A Textbook of Psychology*, W. B. Saunders, Philidelphia, (1958).

[16] E. Kubinyi, Á. Miklósi, F. Kaplan, M. Gácsi, J. Topál, V. Csányi, 'Social Behaviour of Dogs Encountering AIBO, an Animal-Like Robot in a Neutral and in a Feeding Situation', *Behavioural Processes*, *65*, 231-239, (2004).

[17] M. Mori, 'Bukimi No Tani (The Uncanny Valley)', *Energy*, *7(4)*, 33-35, (1970).

[18] H. Brenton, M. Gillies, D. Ballin, D. Chattin, 'The Uncanny Valley: Does it exist?', *Proceedings of Conference of Human Computer Interaction, Workshop on Human animated Character Interaction*, Napier University, Edinburgh, , (2005).

[19] K. F. MacDorman, 'Androids as an experimental apparatus: Why is there an uncanny valley and can we exploit it?', *Proceedings Of the CogSci 2005 Workshop: Toward Social Mechanisms of Android Science*, Stresa, Italy, 106-118, (2005).

[20] H. Hediger, 'The Evolution of Territorial Behavior', in S. L. Washburn (Ed.), *Social Life of Early Man*, Butterworth and Co., London , 34-57, (1961).

[21] J. K. Burgoon, S. B. & Jones, 'Toward a Theory of Personal Space Expectations and Their Violations', *Human Communication Research*, *2(2)*, 131-146, (1976).

[22] S. D. Benford, L. E. Fahlén, 'A Spatial Model of Interaction in Virtual Environments', *Proceedings of the Third European Conference on Computer-Supported Cooperative Work (ECSCW 93)*, Milano, Italy, (1993).

[23] J. N. Bailenson, J. Blascovich, A. C. Beall & J. M. Loomis, 'Interpersonal Distance in Immersive Virtual Environments', *Personality and Social Psychology Bulletin*, *29(7)*, 819-833, (2003).

[24] C. Breazeal, B. Scassellati, 'Infant-like Social Interactions Between a Robot and a Human Caretaker', *Adaptive Behavior*, *8(1)*, 49-74, (2000).

[25] T. Nomura, T. Shintani, K. Fujii & K. Hokabe, 'Experimental Investigations of Relationships Between Anxiety, Nehative Attitudes, and Allowable Distance of Robots (HCI 07)', *Proceedings of the Second Interantioanal Conference on Human-Computer Interation (HCI 07)*, Beijing, P.R. China, 13-18, (2007).

[26] H. Hűttenrauch, K. Severinson Eklundh, A. Green, E. A. Topp, 'Investigating Spatial Relationships in Human-Robot Interaction', *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 06)*, Bejing, China, 5052-5059, (2006).

[27] M. L. Walters, K. Dautenhahn, K. L. Koay, C. Kaouri, R. te Boekhorst, C. L. Nehaniv, I. Werry, D. Lee, 'Close Encounters: Spatial Distances Between People and a Robot of Mechanistic Appearance', *Proceedings of IEEE-RAS International Conference on Humanoid Robots (Humanoids2005)*, Tsukuba, Japan, 450-455, (2005).

[28] M. L. Walters, K. Dautenhahn, K. L. Koay, C. Kaouri, S. N. Woods, C. L. Nehaniv, R. te Boekhorst, D. Lee, I. Werry, 'The Influence of Subjects' Personality Traits on Predicting Comfortable Human-Robot Approach Distances ', *Proceedings of Cog Sci 2005 Workshop: Toward Social Mechanisms of Android Science*, Stresa, Italy, 29-37, (2005).

[29] M. L. Walters, K. Dautenhahn, S. N. Woods, K. L. Koay, R. te Boekhorst, D. Lee, 'Exploratory Studies on Social Spaces between Humans and a Mechanical-looking Robot', *Journal of Connection Science, Special Issue on Android Science*, *18*, 429-442, (2006).

[30] M. L. Walters, K. L. Koay, K. Dautenhahn, R. te Boekhorst & D. S. Syrdal, 'Human Approach Distances to a Mechanical-Looking Robot with Different Robot Voice Styles', *Proceedings of the 17th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2008)*, Munich, Germany, 707-712, (2008).

[31] Koay, K. L., Syrdal, D. S., M. L. . Walters, K. Dautenhahn, 'Living with Robots: Investigating the Habituation Effect in Participants' Preferences During a Longitudinal Human-Robot Interaction Study', *Proceedings of the 16th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2007)*, South Korea, 564-569, (2007).

[32] M. L. Walters, D. S. Syrdal, K. Dautenhahn, R. te Boekhorst, K. L. Koay, 'Avoiding the Uncanny Valley – Robot Appearance, Personality and Consistency of Behavior in an Attention-Seeking Home Scenario for a Robot Companion', *Journal of Autonomous Robots*, *24(2)*, 159-178, (2008).

[33] M. L. Walters, S. N. Woods, K. L. Koay, K. Dautenhahn, 'Practical and Methodological Challenges in Designing and Conducting Human-Robot Interaction Studies', *Proceedings of the AISB'05 Symposium on Robot Companions Hard Problems and Open Challenges in Human-Robot Interaction*, UK, 110-119, (2005).

[34] M. L. Walters, *The Design Space for Robot Appearance and Behaviour for Social Robot Companions*. Doctoral thesis, University of Hertfordshire, http://hdl.handle.net/2299/1806 (2008).

[35] K. L. Koay, M. L. Walters, S. N. Woods, Dautenhahn K., 'Empirical Results from Using a Comfort Level Device in Human-Robot Interaction Studies', *Proceedings of ACM SIGCHI/SIGART 2nd Conference on Human Robot Interaction (HRI '06)*, Salt Lake City, Utah, USA, 194-201, (2006).

[36] S. McCloud, *Understanding Comics: The Invisible Art.*, Harper Collins Publishers, (1993).

[37] K. Dautenhahn, 'Design spaces and niche spaces of believable social robots', *Proceedings of the 11th Annual International Workshop on Robot and Human Interactive Communication (RO-MAN 02)*, Berlin, Germany, 192-197, (2002).

[38] M. L. Walters, D. S. Syrdal, K. L. Koay, K. Dautenhahn & R. te Boekhorst, 'Human Preferences and Perceptions of Robot Appearances', *Proceedings of AISB 09, HRI Symposium*, Heriot Watt Univ., Edinburgh, Scotland, (2008).

[39] T. Nomura, T. Kanda, T. Suzuki, 'Experimental Investigation into Influence of Negative Attitudes Towards Robots on Human-Robot Interaction', *AI & Soc.*, *20*, 138–150, (2006).

[40] M. Heerink, B. Kröse, V. Evers & B. Wielinga, 'The Influence of Social Presence on Acceptance of a Companion Robot by Older People', *Journal of Physical Agents*, *2(2)*, 33-40, (2008).

[41] N. Mitsunaga, C. Smith, T. Kanda, H. Ishiguro & N. Hagita, 'Robot Behavior Adaption for Human-Robot Interaction Based on Policy Gradient Rienforcement Learning', *Proceedings of IROS 05*, 218- 225, (2005).

[42] L. A. Zadeh, 'Fuzzy algorithms', *Information and Control*, *5*, 94-102, (1968).

[43] E. Cox, *The Fuzzy Systems Handbook*, Academic Press Inc., U.S., (1994).

[44] K. L. Koay, Z. Zivkovic, K. Dautenhahn, M. L. Walters, N. Otero, Alissandrakis A., 'Methodological Issues of Annotating Vision Sensor Data Using Subjects Own Judgement of Comfort in a Robot Human Following Experiment', *Proceedings of The 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06)*, Hatfield, UK, 66-73, (2006).

[45] K. Dautenhahn, 'Socially Intelligent Robots: Dimensions of Human - Robot Interaction', *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362(1480)*, 679-804, (2007).

Proceedings New Frontiers in Human-Robot Interaction, K. Dautenhahn (Ed.), symposium at the AISB09 convention, 8-9 April, Edinburgh, Scotland

Addressing User Experience and Societal Impact in a User Study with a Humanoid Robot

Astrid Weiss[1], Regina Bernhaupt[2], Manfred Tscheligi[1] and Eiichi Yoshida[3]

Abstract. Using a preliminary exploratory case study the presented work investigates the feasibility of methods for the evaluation of user experience factors in human-humanoid interaction as well as measurements of societal impact in user studies. The case study is based on two tasks participants had to perform with the HRP-2 robot. The robot was controlled by the participants via speech commands to pick up (task1) and put down (task2) an object onto another place. Main goal of the case study was to explore the methodological concept on how to measure novice users' experiences during the collaboration with the humanoid robot HRP-2 via speech commands and if the general attitude towards robotics changes because of the interaction with the robot. To address users' experiences during the user study, participants were asked after each task to state their thoughts and feelings they had during the interaction (retrospective think aloud). Furthermore, participants were interviewed by means of two validated standardized questionnaires (NARS and AttrakDiff) and an especially developed questionnaire. The preliminary results show that retrospective think aloud is a good way to gather qualitative data on users' experiences. Furthermore a final interview on societal impact of humanoid robots gave insights into how novice users imagine a future society with robots.

1 INTRODUCTION

"User Experience" (abbr. UX) is a concept in the area of Human-Computer Interaction (HCI) referring to the quality of the experience a person has when interacting with a specific design [6] and technology in general [10]. Research on UX currently has three facets [19]: First, UX is seen as a concept that goes beyond the instrumental, e.g. the standard measurements of usability like efficiency and effectiveness have to be extended with a concept like fun. Second, UX is related to emotion and affect referring to the positive aspects of an experience and how to support them (shifting from the perspective of trying to avoid negative consequences). Third, UX represents a concept that is related to a product and user experiences are embedded in a specific situation. Interaction goals, intra-psychological dispositions, the environment, involved people and the product itself have a significant impact on how users experience the product. Each experience has a discrete beginning and an ending. During this time span a user experiences emotions, which are heavily influenced by the introduced components.

There is an increased interest in Human-Robot Interaction (HRI) to establish a positive experience for humans interacting with (humanoid) robots, especially in home settings [28]. Focusing on humans interacting with robots, a possible definition of the user's experience can be based on Alben's [17] general definition that UX includes "aspects of how people use an interactive product: the way it feels like in their hands, how well they understand how it works, how they feel about it while they're using it, how well it serves their purposes, and how well it fits into the entire context in which they are using it".

UX in HRI currently is evaluated using methodologies from the area of HCI. The applicability of these methods is still unclear as the validation of UX evaluation methods is difficult. Especially when interacting with a robot, individual user experiences might be heavily influenced by the individual's general attitude and the overall societal opinion.

Our general goal is to understand to what extent user experience is related to the overall acceptance of robots in society (societal impact). A first step is to investigate a methodological mix (as used in the area of HCI) to understand if HCI methods can be fruitfully used to understand user's experiences when interacting with a robot. The case study thus addressed the thoughts and feelings of novice users when interacting with the HRP-2 robot via speech commands. As the societal impact of robotic agents in future societies can only be roughly predicted, it is of clear interest how novice users themselves imagine their usage of robotic systems. In social sciences the validation of a methodological mix is typically conducted by performing a pre-study with a limited number of participants. These pre-studies can help understand limitations and shortcomings of the method as well as possible influencing factors of an experimental setting.

The pre-study is presented as follows: first related work is discussed, second we present our set-up and the results from a first exploratory study, third we discuss lessons learned for the proposed evaluation set-up.

2 EVALUATING USER EXPERIENCE AND SOCIETAL IMPACT

Several evaluation studies on perception of robots in collaborative settings have been conducted. Hinds [26] e.g. investigated people's perception of robots as working partners. She investigated differences in reliability and feeling of responsibility on the part of the human and the robot. Results showed that participants relied more on humans as working partners, but interestingly the results were different for the concept of responsibility. There was slightly no difference between participants' feelings of responsibility for a human and a robotic working colleague. Also the comparison between human-like and machine-like robots in terms of the willingness to rely on was not significant.

1 HCI&Usability Unit, ICT&S Center, University of Salzburg, Sigmund-Haffner-Gasse 18, 5020 Salzburg, Austria. Email: {firstname.lastname}@sbg.ac.at.

2 IHCS-IRIT, University Paul Sabtier, 118 Route de Narbonne, 31062 Toulouse Cedex 9. Email: regina.bernhaupt@irit.fr

3 National Institute of Advanced Industrial Science and Technology (AIST), Ibaraki 305-8564, Japan; Email: yoshida@aist.go.jp;

Another interesting aspect of the integration of robots into working environments was revealed by [3]. Based on a long-term ethnographic study, where a delivery robot was introduced into the everyday working environment of a hospital they could show that perception of the robot depends on the user group. Their results showed significant differences on how people responded to the robots in two different organization units in the hospital. The delivery robot was differently integrated into the work-flow of the two units and also the way of interaction differed.

Some aspects of UX in HRI are extracted from avatar research since many research objectives from UX in HRI are also pursued by scientists of ECAR (Embodied Conversational Agent Research) (e.g. the robot serves as a representation of a human, or robots are used to investigate embodied models of social behaviour [28]). The embodiment (or morphology) [28] has an impact on how a robot is perceived. It also implies capabilities of a system, which in some cases cannot be fulfilled. If a robot is designed biologically inspired with anthropomorphic features, then users will expect human-like capabilities. In some cases users may experience fear as robots are designed too realistic. It is important for people that there is a clear difference between humans and robots [27]. In 1970 Masahiro Mori, proposed the well known hypothetical graph of the uncanny valley, which predicted that the more human a robot looks, the more familiar it is, until a point is reached at which subtle imperfections make the robot seem eerie [15].

By incorporating existent navigation technique metaphors in HRI, a more positive UX is obtained. This can be achieved by transferring navigation strategies and interface designs of computer games (first person shooters) to robot navigation [25].

However, the evaluation of user experiences while interacting with a robot is only sparsely reflected in the current literature. Syrdal et al. investigated how context factors in HRI video prototyping influence the user's experience and attitude, by means of qualitative interviews [7]. However, to our knowledge there is no work directly referring to the evaluation of UX in terms of pragmatic and hedonic quality. Other evaluations refer to factors that are related to the concept of UX, like acceptance, see e.g. [24][21][23][22]. Attitudes [18], especially negative attitudes [29]; [5], have been investigated in detail. An excellent overview on existing measures for single UX factors in HRI can be found in [4]. However, studies on "general" user experience have not been conducted so far.

Societal impact on the other hand was assessed with a variety of methods. Sakamoto and Ono [8] used a lab-based approach to investigate if in the future robots will construct or collapse human relations. They conducted an experiment based on a communication situation between two humans and one robot with three conditions: (1) robot agrees with humans, (2) robot disagrees with humans, (3) robot agrees only with one human. However, the results of this one laboratory-based experiment could not show a clear tendency if human relations will collapse or not in future societies due to the introduction of robots.

Large questionnaire based surveys have been conducted to get insights on people's expectations towards the future society and the societal impact of robots. Arras and Cerqui [16] conducted a paper and pencil based survey with 2042 participants in conjunction with the "Robotics" pavilion at the Swiss National Exhibition Expo.02. The results testified a positive attitude towards potential robotic co-workers, robots as flat-mates, and even for robotic body parts. The authors stress on the one hand that context-related surveys with a sufficient number of participants give valuable insights on societal trends, but on the other hand they admit that the conducted survey is only valid for Europe and should be conducted in the USA and Asia as well to get more generalizable results.

The relationship of usability, user experience, (social) acceptance and societal impact is currently investigated in the Robot@CWE project funded by the 6th EU IST framework. Weiss et al. [2] showed how breaching experiments can help to understand social acceptance and societal impact factors. The following case study lays the basis for the evaluation of user experience and societal impact.

3. THE CASE STUDY

3.1 Methodological Considerations

The case study was based on two tasks, where participants had to control the HRP-2 robot via speech commands to pick up (task 1) and put down an object on another place (task 2). Four participants took part in this preliminary case study (1 female, 3 males). The youngest participant was aged 23, the oldest one 39, with an average age of 27.5 years.

The content-related research questions of this user study were:

- How do novice users experience the interaction with the humanoid robot HRP-2 when interacting via speech commands?
- How do users perceive the system in terms of usability?
- Does the general attitude towards robotics change because of the interaction with the robot?
- How do people imagine the future society after interacting with the robot?

From a methodological view point the focus of interest was to verify if the standardized questionnaires from other disciplines, as well as the first self-developed questionnaire on UX indicators are reasonable to evaluate the proposed factors. Clearly a statistical validation of the questionnaire data was not possible with four participants, but the combination of different questionnaires and the qualitative data showed tendencies which study settings and instruments can be reasonably used. Furthermore, it was necessary to investigate the workload of such a study setting for the participants and which methodological considerations have to be taken into account for future experiments.

3.2 Study Setting

The user study with HRP-2 took place at the LAAS - CNRS institute in Toulouse from 8th to 11th of July 2008. It was based on two tasks which participants had to conduct with the humanoid HRP-2 robot. 15 clones of the HRP-2 robot exist. University of Tokyo and AIST conduct research with these robotic platforms on robot autonomy. Currently the Joint French-

Japanese Robotics Laboratory (JRL) in collaboration with LAAS-CNRS and AIST are working on enabling HRP-2 understanding and executing commands like "give me the orange ball", as autonomously as possible (details on technical constraints of this work can found in [13]). This built the basis for the presented case study, which was conducted with HRP-2 number 14.

The study design was based on two tasks (see table 1) participants had to complete by giving HRP-2 speech commands More details about the implementation of the two tasks can be found in [9]

Table 1: The two Tasks

Task 1	Task 2
Instruction: "Tell the robot to pick up the orange ball. Therefore, you have to use the following commands (after each command, wait for the reaction/answer of the robot, before you start with the next one)"	Instruction: "Your task is now to tell the robot to put the orange ball on the table. Therefore, you have the following commands"
Commands needed: "14" (is the general command which activates the robot, and its name) "Go to the green box" "Look down" "Take the orange ball"	Commands needed: "14" (is the general command which activates the robot, and its name) "Turn to the left" "Go to the yellow table" "Look down" "Put the orange ball on the yellow table"
Additional commands: Turn to the left Turn to the right	Additional commands: Turn to the left Turn to the right

The tasks were introduced to the participant by the experimenter reading aloud the following scenario:
``*Imagine you are working at a construction site and need a special tool. A humanoid robot, its name is fourteen, is supporting you by providing you with your needed tools. You can control the robot with predefined commands, therefore it picks-up and transports the tool''*

Afterwards the participant received a microphone to give the robot the speech commands. HRP-2 executed the commands autonomously, therefore the additional commands were necessary, as it could happen that due to wrong speech recognition the robot executed motions that were not expected by the participant (e.g., the robot picks up the purple ball instead of the orange one).

Figure 1 shows the setting, with the HRP-2 robot (the execution of the motion was observed by a researcher in the background due to safety reasons), the participant with the microphone, the experimenter taking observational notes, and the green box, with a purple and an orange ball.

Figure 1: Study Setting

Four participants took part in the case study, one was female three were male and they were between 20 and 40 years old. The requirements to select the participants were: (1) know English well (the study was conducted in English); (2) not have any pre-experience with robots (as insights should be gathered how novice users experience the interaction)

3.3 Instruments and Measurements

3.3.1 Usability Measurements

The whole user study was video-taped. Task completion and task duration were recorded during the user study to address the usability indicators effectiveness and efficiency. Furthermore, a "retrospective think aloud" was conducted. After each task, participants were asked about improvements and changes they would suggest to make the interaction with HRP-2 easier in future.

Additionally the system usability scale questionnaire (SUS) was used to address the perceived usability of the interaction. This questionnaire was developed at Digital Equipment Corp. as a tool which allows a simple and quick standardized evaluation of system usability. The scale consists of ten statement-based items, giving a global view of subjective assessments of usability. The statements are rated by the participants on a five-point Likert scale from "strongly disagree" to "strongly agree". The overall scale takes into account the effectiveness, efficiency, and satisfaction of participants with a system. Thus the result of the questionnaire is a single number representing a composite measure of the overall usability of the system being studied, meaning that scores for individual items are not meaningful on their own. The range for this cumulative score is in between 0-100: "80-100: users like the system"; "60-79: users accept the system"; "0-59: users do not like the system".

3.3.2 User Experience Measurements

To get general insight on UX when interacting with the HRP-2 robot, the validated standardized AttrakDiff questionnaire was used. This questionnaire was designed to measure beyond the usability (pragmatic quality) also the user experience (hedonic quality) of a system [20]. The questionnaire is a semantic differential consisting of numerous antithetic word-pairs, e.g. "disagreeable – likable". All items have to be graduated by the participants on a scale ranging from the negative word pole to the positive word pole; the negative pole (in sense of the scale) is coded with -3 and the positive with 3.

The results are presented in four different sub scales: "PQ = pragmatic quality of the system", "HQ-I = Hedonic Quality Identification", "HQ-S = Hedonic Quality Stimulation", and "ATT = Attractiveness". The pragmatic quality describes the usability of the system and indicates how successful users are in achieving their goals using the product. The hedonic quality - identity scale describes to what extent the system allows the user to identify with it, whereas the hedonic quality - stimulation is an indicator to what extent the system can support stimulating functions, contents, and interaction- and presentation-styles. Finally attractiveness describes a global value of how the quality of the system is perceived by participants. The hedonic and pragmatic qualities are independent of one another, and contribute equally to the rating of attractiveness.

Furthermore, a questionnaire was developed (subsequently called UX-questionnaire) which tried to address five predefined UX factors with five statements each and had to be rated on a 7-point Likert scale (where 7 was the best rating in sense of the scale) by the participants.

The five factors were chosen based on an extensive literature review: Emotion [28], Embodiment [28], Feeling of Security [12], Human-oriented Perception [28], Co-Experience [14] and are part of the USUS evaluation framework [1]. Table 2 presents all items of the questionnaire with the corresponding factors.

Table 2

Statements addressing User Experience Factors

Statement	Factor*
I liked the size of the robot.	Emb
Interacting with the robot is fun.	E
When talking to the robot, I feel like talking to a human.	Co
I am happy when the robot understands my commands.	E
I think that the robot is vulnerable to hackers.	FoS
I can interact with the robot like I interact with other humans.	Co
I am disappointed if the robot does not understand my commands.	E
I liked that the robot looked similar to a human.	Emb
I hesitate to use the robot for fear of making errors that will harm me.	FoS
When working with the robot I perceive it as working in a team.	Co
I perceive the robot as a social actor	HoP
I liked that the robot has human like features: face, ears, eyes, etc.	Emb
I feat to use the robot, as an error might harm the robot.	FoS
I liked that the robot detected my face.	HoP
I feel good when interacting with the robot.	Co
I liked the physical co-location of the robot.	Emb
I perceive that the robot is intelligent	HoP
I am angry if the robot does not understand my commands	E
I liked the design of the robot.	Emb
I enjoyed talking with the robot	HoP
The robot could become a companion for me.	Co
I feel secure when working with the robot.	FoS
I felt afraid of the robot.	E
I liked that the robot understands my voice commands	HoP
I perceive the robot as safe.	FoS

*Emb: Embodiment; E: Emotion; Co: Co-Experience; FoS; Feeling of Security; HoP: Human-oriented Perception;

To measure if the participant's attitude towards humanoid robots as working colleagues changes during the user study the NARS questionnaire was used.

This questionnaire is based on a psychological scale to measure the negative attitudes of humans against robots and is originally developed by [29]. In the described user study the questionnaire was used in an English version [5]. This questionnaire tries to visualize which factors prevent individuals from interacting with robots. The questionnaire consists of 14 questions, which have to be rated on a 5 point Likert scale ranging from "Strongly Disagree" to "Strongly Agree". The 14 questions build three sub scales: S1 = Negative attitude toward situations of interaction with robots; S2 = negative attitude toward social influence of robots; S3 = negative attitude toward emotions in interaction with robots.

3.3.3 Societal Impact Measurements

To address participants' assessment of the future society regarding humanoid robotic working colleagues, an in-depth interview was conducted. The interview was based on four questions which were discussed with the participants.

- In which way could robots be integrated into working life in future?
- How could life change if robots are integrated into a construction site?
- How will the usage of robots in the working context influence the future education system?
- How could you imagine that society will support the use of robots in the future?

3.4 Procedure

The study started in a room separated from the test-setting, where participants had to fill in the NARS at first. Then the two tasks were conducted. After each task participants were asked what they thought and felt during the interaction with the robot, if they had any problems during the interaction, and if they would like to change or improve something to make the task fulfillment easier. Then participants rated the difficulty of the task. After the two tasks were carried out, participants filled in the questionnaires in the following order:

SUS

AttrakDiff

UX-Questionnaire

NARS (second round)

After the questionnaires were filled in the user study was concluded with the in-depth interview on participants' imagination about how future society could look like regarding the integration of robotic technology.

5 RESULTS

5.1Insights on Perceived Usability

Table 3 gives an overview on the task completion of all participants for all tasks. The results show that participants were very successful in solving the tasks, and thus did not experience the interaction with HRP-2 as very difficult.

Table 3: Task Completion

	Task 1	Task 2
Solved successfully	2	3
Solved with help	2	1
Not solved	0	0
Average rating of difficulty[4]	1.5	1.33

The tables 4 and 5 give an overview on the clustered comments participants stated in the retrospective think aloud. The number in brackets is the number of the participant (not a count!).

Table 4: Retrospective think aloud – Task 1

Participants liked	Participants disliked
Robot answered in an nice way (3)	Speed of the robots reaction (1, 2, 3, 4) (too slow)
No need of additional and repeated commands (3)	TP felt insecure if the robot is close enough to grasp the ball and therefore already had to look on how to proceed with the next command. (3)
Robot describes all moves very good and helpful (3)	There was orange and purple misunderstanding (1)
Impressed, really impressed, if it was doing something wrong it was my fault (2)	I did not understand what it said, I had just to give commands (1)
Amazing, like in a movie (3)	Not every move needs a command (4)
The fears the robot would not understand me, were not realized.	When it still needs human it is not useful (4)
Feeling very close to the process (1)	

Table 5: Retrospective think aloud – Task 2

Participants liked	Participants disliked
Easy to use (2, 3)	Difficult: I had to figure out and understand how it works in term machine (1)
I thought I need to talk slow, I would not have thought it is so precise (2)	I saw the robot as a machine, not an human (1)
Cool and easy to do (3)	The robot is not human like, I do not remember that the robot detected my face (4)
I was comfortable when the robot explained why he did not do the command (3)	Comprehension should be better (orange-purple) (2)
	I thought he will remind the yellow table that would be more logical (3)
	It should remind things then it gets easier (3)
	I was a little bit stressed as it did not understand the first command.
	Its movement is slow (4)
	The time between answer and movement is too long (4)
	Improve: having different programs, one for talking, one for movement (4)
	Lack of pronunciation (p of put, and p of purple)

[4] from 1 “very easy” to 5 “very difficult”

The SUS questionnaire equaled a surprisingly high result of 78.8, meaning that the participants accepted the HRP-2 robot in terms of its usability. However the SUS is no absolute ratio, it depends on the experimental context; in this case study participants evaluated the speech interaction with the robot.

5.2 Insights on User Experience

The results for the AttrakDiff questionnaire (see table 6) presented a neutral image of the general UX when interacting with the robot. The comparison of the mean values for each scale shows that people experienced the robot rather neutral (all mean values are around 0). This means that there is still room for improvement of the user experience of the robot in terms of “Pragmatic Quality”, “Hedonic Quality Identification”, “Hedonic Quality Stimulation”, and “Attractiveness”.

Table 6: Results AttrakDiff

	PQ	HQI	HQS	ATT
Mean	0.0000	-.5357	.0357	.0714
S.D.	.37796	.44224	.62133	.24744

For the UX-Questionnaire the UX factors were calculated by summing up the ratings of the participants (see table 7). They can get values from 1 to 7. The higher the value is, the more positively the participants connected the form of experience with the tested robot.

Table 7: Results UX Questionnaire

UX-Factors	min	max	mean	SD
Co-experience	3.20	4.40	3.7000	.25166
Embodiment	4.60	6.20	5.3333	.46667
Emotion	5.20	6.60	5.8500	.29861
Human-Oriented Perception	4.20	5.80	4.9333	.46667
Feeling of Security	4.20	5.80	5.2000	.38297

The factor “emotion” comes off very well with an average value of 5.85. Also “embodiment”, “feeling of security” and “human-oriented perception” have quite good values with 5.33, 4.93 and 5.2. Only located in the middle of the field is the factor “co-experience” with an average value of 3.7.

The NARS questionnaire revealed to most impressive result (see table8). There is a significant difference between the rating of “negative attitude toward situations of the interactions with robots” (S 1) before and after the test. After the experiment/test the rating of the “negative attitude toward situations of the interactions with robots” is significantly lower than before the experiment. Also the rating of “Negative Attitude toward Social Influence of Robots” (S 2) has significantly changed. Before the test the rating of “Negative Attitude toward Social Influence of Robots” has been significantly higher than after the test. The rating of “Negative Attitude toward Emotions in Interaction with robots” has also decreased, but not significantly.

Table 8: Results NARS Questionnaire

		Mean	SD	t	sig.
Nars S1	t1	11.75	4.11	5.196	.014
	t2	10.25	3.95		
Nars S2	t1	14.00	2.16	3.434	.041
	t2	10.75	2.36		
Nars S3	t1	9.75	2.06	-1.507	.229
	t2	7.50	2.38		

5.3 Insights on Perceived Societal Impact

When discussing about the question in which way robots could be integrated into future working life, 3 out of 4 participants use the term “replacement”, which implicates that if robots are integrated into working life, automatically something or someone other is excluded from work (e.g. humans, machines). Therefore “replacement” is a negative term.

Basically the participants are of the opinion that robots can replace on the one hand “slavery work” and on the other hand dangerous work. The term “slavery work” is defined as both repetitive tasks in construction factories and domestic work like cleaning, cooking, washing the dishes, cutting the grass or even driving a car. As examples for dangerous work the participants name “finding lost people in the mountains” or “lifting heavy things”. In the participants` opinion the advantage of using robots is that they can be replaced when they are broken, the disadvantage is that human life is nowadays cheaper than robots (e.g. in China).

The following citations show that the participants are thinking differently about the integration of robots into the working life.

- “It is impressive, but dangerous, we are a human society not a robot society, we have to keep humanity” (participant 2)
- “Environments working harmful for humans” (participant 1)
- “Robots are useful as assistance” (participant 3)
- “You can replace a robot if it is broken” (participant 4)

Two participants foreground the dangers aroused through the integration of robots into working life (e.g. unemployment), the other two participants do not mention dangers but give priorities to the advantages that could accrue by the usage of robots in the working life. So half of the participants vision the integration of robots into the working life positively, half of the participants negatively.

Concerning the question how life could change if robots are integrated into a construction area, the participants forecast both positive and negative effects. One negative effect is the unemployment as a result of the replacement of humans by robots. Also negatively seen is the competition between robots and humans. The participants are of the opinion that robots will compete because they do not need to eat or drink and therefore are more reliable, meaning that the participants think that the costs of the workers are the most crucial employment-factor. It is seen positively that there will be less accidents in future working life with robots, so work will be safer, there will be less health problems and people will have more spare time where they have the possibility to do funny things while robots are doing their work (e.g. their household). Also positively noted is that work that only can be done by humans will be more valued in a future working life with robots.

Subsequently it can be said that participants think there will be more unemployed people, because robots are doing their work. At the same time people will have more time for doing funny things. What the participants do not mention is the inconsistence between increasing unemployment and increasing fun in the spare time. There will be probably increasing fun for one group of people and less fun for the other unemployed group that cannot afford robots for domestic work.

Concerning the question how usage of robots in the working context will influence the future education system, participants have two different concepts. One part of participants is of the opinion that people have to learn the usage of robots in school (e.g. school subject “technology” beginning at the age of about 15). The other part of participants thinks that children will be able to learn outside school how to use robots (e.g. “parents will teach the children the dangers of robots, it´s like with cars”).

Therefore, we can distinguish between two robot-education-types:

- Type 1: A robot is only one more technology, and there is no need for special education. Computers in education are enough.
- Type 2: There should be an introduction of robots in school, so people get to know how to interact with a robot.

The last question was about how the society would support the use of robots in the future. For participants the usage of robots in future is mostly a benefit-cost-question. They think it depends on the price, the tasks the robot is able to do, the speed it is fulfilling the tasks with and so on. But also the ease of use and the main aim will be important factors. (If robots are only dedicated to economic growth they will not be as accepted as if robots that are dedicated to our life). Participants are of the opinion that if people get used to robots (if they are comfortable), they will be just another technology, like cars. For the young generation robots could be domestics, like computers, especially if they get cheaper and quicker.

6 LESSONS LEARNED FOR THE PROPOSED EVALUATION SET-UP

Considering our experiences we recognized the following issues as being crucial to successfully address UX and societal impact in a user study with a humanoid robot:

1. Participant sampling and requirements: When conducting a user study in HRI with novice users it can be valuable to consider the educational background. In the retrospective think aloud and in the societal impact interview participants often underlined their assumptions based on their education and work background: “As I am working in education; I can tell you that robots should not teach children”; “As I studied computer science, I have a mental model how the robot vision works”.
2. Method mix: It is recommended to combine qualitative and quantitative methods to gain a more holistic picture of UX. Linking the data from the pre-structured questionnaires with the retrospective think aloud and the societal impact interviews facilitated the clustering and interpretation of the gathered data. Furthermore we could show that the proposed HCI questionnaires can be fruitfully used to understand user’s experiences when interacting with a humanoid robot.
3. Study setting: The UX factor feeling of security could not be reasonably addressed in the pre-study setting where a researcher had to observe the motions of the robot. Three participants stated difficulties in answering these questions of the UX-questionnaire: “I think the robot cannot be safe, as you have to watch it all the time”. Thus, it is recommended to assess the factor feeling of security in controlled settings that do not bias the answer behavior of the participants.
4. Procedure: All participants mentioned positively that they had the possibility to fill in the NARS questionnaire a second time: “That’s good, because my attitude actually changed”.

7 CONCLUSIONS AND FUTURE WORK

The main goal of this case study was to explore how novice users experience the collaboration with the humanoid robot HRP-2 when interacting via speech commands and if the general attitude towards robotics changes because of the interaction with the robot.

To investigate the feasibility of the study set-up and the methodological approach, we conducted a first case study with four participants. Our intention was to get a better understanding of first time user reactions on humanoid robots and to find out if these reactions influence participants’ attitude towards robots. The context of a user study where participants can directly control the robot via speech commands in combination with several questionnaires and an in-depth interview have proven its value, as different UX factors of the participants could be addressed and also discussed on the reflective level with the participants (retrospective think aloud and final interview). Furthermore, the workload seemed reasonable for the participants and built the basis for three similar future experiments.

To validate the questionnaire which was based on five UX factors in term of the scales, a summative analysis of the data gathered in this case study combined with the three future user studies and a broad online survey (we are expecting a total of 500 participants) is planned.

ACKNOWLEDGEMENTS

This case study was conducted in the framework of the ROBOT@CWE project (funded by FP6-2005- IST-5). The authors like to thank all partners from the project and gratefully acknowledge the collaboration with the researchers from LAAS-CNRS, with special thanks to Anthony Mallet.

REFERENCES

[1] A. Weiss, R. Bernhaupt, M. Lankes, M. Tscheligi, ‘The usus evaluation framework for human-robot interaction’, *in AISB2009: Proceedings of the Symposium on New Frontiers in Human-Robot Interaction* (2009).

[2] A. Weiss, R. Bernhaupt, M. Tscheligi, D. Wollherr, K. Kuehnlenz, and M. A Buss, ‘Methodological variation for acceptance evaluation of human-robot interaction in public places’, in *IEEE RO-MAN 2008: International Symposium on Robot and Human Interactive Communication, Munich, Germany*, (2008).

[3] B. Mutlu and J. Forlizzi, ‘Robots in organizations: the role of workflow, social, and environmental factors in human-robot interaction’, in *HRI ’08: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pp. 287–294, New York, NY, USA, (2008). ACM

[4] C. Bartneck, E. Croft, and D. Kulic, ‘Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots’, International Journal of Social Robotics, (2009).

[5] C. Bartneck, T.Nomura, T. Kanda, T. Suzuki, and K. Kato, ‘A cross-cultural study on attitudes towards robots’, in *HCI International*, Las Vegas, USA, (2005).

[6] D. Knemeyer and E. Svoboda, ‘User experience - ux’, *interactiondesign. org*, (2005).

[7] D. S Syrdal, N. Otero, and K. Dautenhahn, ‘Video prototyping in human-robot interaction: results from a qualitative study’ *15th European Conference on Cognitive Ergonomics: the Ergonomics of Cool interaction*, ACM, New York, NY, 1-8 (2008).

[8] D. Sakamoto, and T. Ono, 2006. Sociality of robots: do robots construct or collapse human relations?. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot interaction*

(Salt Lake City, Utah, USA, March 02 - 03, 2006). HRI '06. ACM, New York, NY, 355-356.

[9] E. Yoshida, A. Mallet, F. Lamiraux, O. Kanoun, O. Stasse, M. Poirier, PF. Dominey, J-P. Laumond, and K. Yokoi, "'give me the purple ball" – he said to HRP-2 no.14', in Proc. 2006 IEEE-RAS Int. Conf. on Humanoid Robots, (2007).

[10] J. McCarthy and P. Wright 'Technology as Experience. London : MIT Press (2004).

[11] K. Battarbee, 'Defining co-experience', in DPPI '03: Proceedings of the 2003 international conference on Designing pleasurable products and interfaces, pp. 109–113, New York, NY, USA, (2003). ACM.

[12] K. Dautenhahn, M. L. Walters, S. Woods, K. L. Koay, E.A. Nehaniv, C. L. Sisbot, R. Alami, and T. Simeon, 'How may I serve you?: a robot companion approaching a seated person in a helping context', in *HRI '06: Conference on Human-Robot Interaction*, pp. 172–179, Salt Lake City, Utah, USA, (3 2006). ACM.

[13] K. Kaneko, F. Kanehiro, S. Kajita, H. Hirukawa, T. Kawasaki, K. Akachi M. Hirata, and T. Isozumi, 'The humanoid robot HRP-2', in Proc. 2004 IEEE Int. Conf. on Robotics and Automation, pp. 1083–1090, (2004).

[14] K. Severinson-Eklundh, Anders Green, and Helge Hüttenrauch, 'Social and collaborative aspects of interaction with a service robot', *Robotics and Autonomous Systems*, **42**(3-4), 223–234, (2003).

[15] K.F. MacDorman, and H. Ishiguro, 'The uncanny advantage of using androids in social and cognitive science research'. Interaction Studies, 7, 3 (2006), 297-337.

[16] K.O. Arras and D. Cerqui, 'Do we want to share our lives and bodies withrobots? a 2000 people survey', Technical report, (2005 2005). Technical report Nr. 0605-001.

[17] L. Alben, 'Quality of experience: defining the criteria for effective interaction design', *interactions*, **3**(3), 11–15, (1996).

[18] L. Takayama, W. Ju, and C. Nass, 'Beyond dirty, dangerous and dull: what everyday people think robots should do', in *HRI '08: 3rd ACM/IEEE international conference on Human robot interaction*, pp. 25–32, New York, NY, USA, (2008). ACM.

[19] M. Hassenzahl and N, Tractinsky, 'User Experience – a research agenda', Behavior and Information Technology, **25**(2), 91-97 (2006).

[20] M. Hassenzahl, 'The thing and I: understanding the relationship between user and product', in *Funology. From Usability to Enjoyment*, eds., M. Blythe, C. Overbeeke, A. F. Monk, and P. C. Wright, 31–42, Kluwer, Dordrecht, (2003).

[21] M. Heerink, and B. Kröse, B. Wielinga, and V. Evers, 'Studying the acceptance of a robotic agent by elderly users', *International Journal of Assistive Robotics and Mechatronics*, **7**(3), 25–35, (9 2006).

[22] M. Heerink, B. Kröse, B. Wielinga, and V. Evers, 'Enjoyment intention to use and actual use of a conversational robot by elderly people', in *HRI '08: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pp. 113–120, New York, NY, USA, (2008). ACM.

[23] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, 'The influence of a robot's social abilities on acceptance by elderly users', in *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pp. 521–526, (2006).

[24] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, 'The influence of social presence on enjoyment and intention to use of a robot and screen agent by elderly users', in *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, pp. 695–700, (2008).

[25] M. W., Kadous, Sheh, R. K., and C. Sammut, 'Effective user interface design for rescue robotics', HRI '06. 1st ACM SIGCHI/SIGART Conference on Human-Robot interaction (Salt Lake City, Utah, USA, March 02 - 03, 2006), pp. 250-257, New York, NY, (2006). ACM.

[26] P. J. Hinds, T. L. Roberts, and H. Jones, 'Whose job is it anyway? a study of human-robot interaction in a collaborative task', *Human-Computer Interaction*, **19**(1-2), 151–181, (2004).

[27] S. Turkle, 'The second self: Computers and the human spirit'. Simon & Schuster, New York, (1984).

[28] T. Fong, I. Nourbakhsh, and K. Dautenhahn, 'A survey of socially interactive robots', *Robotics and Autonomous Systems*, **42**, 143–166, (3 2003).

[29] T. Nomura, T. Kanda, and T. Suzuki, 'Experimental investigation into influence of negative attitudes toward robots on human-robot interaction', *AI Soc.*, **20**(2), 138–150, (2006).

Proceedings New Frontiers in Human-Robot Interaction, K. Dautenhahn (Ed.), symposium at the AISB09 convention, 8-9 April, Edinburgh, Scotland

The USUS Evaluation Framework for Human-Robot Interaction

Astrid Weiss[1], Regina Bernhaupt[2], Michael Lankes[1] and Manfred Tscheligi[1]

Abstract. To improve the way humans are interacting with robots various factors have to be taken into account. An evaluation framework for Human-Robot Collaboration with humanoid robots addressing usability, social acceptance, user experience, and societal impact (abb. USUS) as evaluation factors is proposed (see figure 1). The theoretical framework USUS is based on a multi-level indicator model to operationalize the evaluation factors. Evaluation factors are described and split up into several indicators, which are extracted and justified by literature review. The theoretical factor indicator framework is then combined with a methodological framework consisting of a mix of methods derived and borrowed from various disciplines (HRI, HCI, psychology, and sociology). The proposed method mix allows addressing all factors within the USUS framework and lays a basis for understanding the interrelationship of the USUS Factors.

1 INTRODUCTION

Integrating humanoid robots into human working environments is a challenging endeavour. It is important to consider that users in human-robot interactions can face severe problems and difficulties. Studies showed that users perceive autonomous robots differently than other computer technologies [32]: Autonomous robots often lead to a far more anthropomorphic mental model than other interface technologies; moreover, as mobile robots always have to adapt to their environment they also have to conform to the humans they are working with, thus the interaction of robots and humans has to be negotiated. Furthermore, robots "learn" about themselves and their world, which heavily distinguishes them from traditional computing technology [32]. All these issues have a very strong influence on the users' work environment, on the way people collaborate, and on the way they experience robotic co-workers.

Such new technologies have a considerable impact on various factors of the interaction between humans and robots: usability, user experience, social acceptance, and societal impact have to be carefully investigated with appropriate measurements and approaches, to lay the basis for future ways of working, including robots that increase productivity and maintain safety. The theoretical and methodological evaluation framework USUS, which was developed from a human-centered HRI perspective [9] for evaluating usability, social acceptance, user experience, and societal impact for working scenarios with humanoid robots, can help us understand how to improve the construction of robots. The framework enables a positive user experience of all users – either individual or in groups, to enhance social acceptance and to support in general positive attitude towards (humanoid) robots in society.

[1] HCI&Usability Unit, ICT&S Center, University of Salzburg, Sigmund-Haffner-Gasse 18, 5020 Salzburg, Austria. Email: {**firstname.lastname**}**@sbg.ac.at**.

[2] IHCS-IRIT, University Paul Sabtier, 118 Route de Narbonne, 31062 Toulouse Cedex 9. Email: regina.bernhaupt@irit.fr

2 STATE OF THE ART

As the research field of Human-Robot Interaction (HRI) is young, but evolving, the need for theoretical and methodological frameworks increases. As Bartneck et al. [4] claim: "If we are to make progress in this field then we must be able to compare the results from different studies". The framework proposed in this position paper should contribute to this aim and therefore take into account efforts already being made in this direction.

Thrun [43] provides one of the first theoretical frameworks for HRI based on the distinction of robots into three different kinds: industrial robots, professional service robots, and personal service robots. He describes in detail the different human-robot interface capabilities, different potential user groups, and the different contexts of use, which lays the first basis for future evaluation approaches in HRI. Thrun himself states in the abstract of this article: "The goal of this article is to introduce the reader to the rich and vibrant field of robotics, in hope of laying out an agenda for future research on human robot interaction" [43].

A similar intention was subject to Yanco et al. [48], who updated their taxonomy of human-robot interaction from 2002, to provide a basis for research in this area. They already address multiple research areas like HCI (Human-Computer Interaction), CSCW (Computer Supported Cooperative Work), and social sciences to offer a holistic picture of research aspects, proposing 11 categories, which need to be considered when investigating the interaction between a human and a robot. The taxonomy of Yanco et al. allows the comparison of different HRI research approaches and therefore is a first step in the direction of making HRI research more generalizable.

However, besides the theoretical frameworks, the need in HRI grows to define metrics to measure the success of robotic systems in a comparable way. This need became obvious with the inaugural workshop on "Metrics for Human-Robot Interaction" held in conjunction with the 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI 2008). The goal of this workshop was "to propose guidelines for the analysis of human-robot experiments and forward a handbook of metrics that would be acceptable to the HRI community and allow researchers both to evaluate their own work and to better assess the progress of others."

A first attempt in this direction was already made by Steinfeld et al. [42]. In their framework they proposed five metrics for task-oriented human-robot interaction with mobile robots: (1) Navigation, (2) Perception, (3) Management, (4) Manipulation, and (5) Social. Furthermore, they mention relevant “biasing effects”, which also have to be taken into consideration when evaluating the proposed metrics: communication factors (like delay, jitter, and bandwidth), robot response (like system lag and update rate), and the user (like training, motivation, and stress). Steinfeld et al. [42] provided above all metrics for usability factors (from an HCI perspective); however they stressed in their conclusion that their proposed “evaluation plan is to provide a living, comprehensive document that future research and development efforts can utilize as a HRI metric toolkit and reference source”.

Bartneck et al. [4] on the other hand tried to provide a standardized toolkit to measure user experience factors in HRI: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety. Based on an extensive literature review they developed five validated questionnaires using semantic differential scales to evaluate human-robot interaction in terms of these factors. “It is our hope that these questionnaires can be used by robot developers to monitor their progress”.

The Dutch research group around Marcel Heerink is focusing on the further development of the UTAUT model (Unified Theory of Acceptance and Use of Technology, see [46] for human-robot interaction in elder care institutions. They investigate which factors have an influence on the intention to use robotic agents. They consider factors like enjoyment [24] social presence [23] and social abilities [22]. Based on that, they want to develop a framework targeting studies on the acceptance of robotic agents by the elderly.

The theoretical and methodological framework proposed in this position paper addresses usability, social acceptance, user experience, and societal impact of humanoid robots used in collaborative tasks and intends to be applied to answer the general question ***if people experience robots as a support for cooperative work and accept them as part of society*** and thus give an holistic view on evaluating humanoid robots. Therefore, the proposed evaluation framework consists of two parts: (1) a theoretical framework defining the relevant evaluation factors and indicators combined with (2) a methodological framework explaining the methodological mix to address these factors during the evaluation of human-robot interaction.

3 THE FACTOR MODEL

The proposed evaluation framework for Human-Robot-Collaboration with humanoid robots is based on a multi-level indicator model targeting the factors *usability, social acceptance, user experience, and societal impact* as evaluation goals. The factors are selected to identify socially acceptable collaborative work scenarios where humanoid robots can be deployed beneficially to convince society to positively support the integration of humanoid robots in a human's working environment. The driving motivation for choosing these factors is to support user-centred evaluation approaches in HRI, going beyond pure usability studies. Although the framework was developed on an intense literature review taking into account existing frameworks and evaluation approaches in HRI, it cannot be guaranteed that is an exhaustive framework.

All factors chosen for evaluation are based on several indicators, which can be addressed with a methodological mix to be assessed during an iterative design process of human-robot collaboration. To justify the selection of factors and indicators several case studies are currently conducted in the framework of the EU-funded FP6 project “Robot@CWE: Advanced robotic systems in future collaborative working environments”. Figure 1 visualizes the combination of the theoretical and methodological framework.

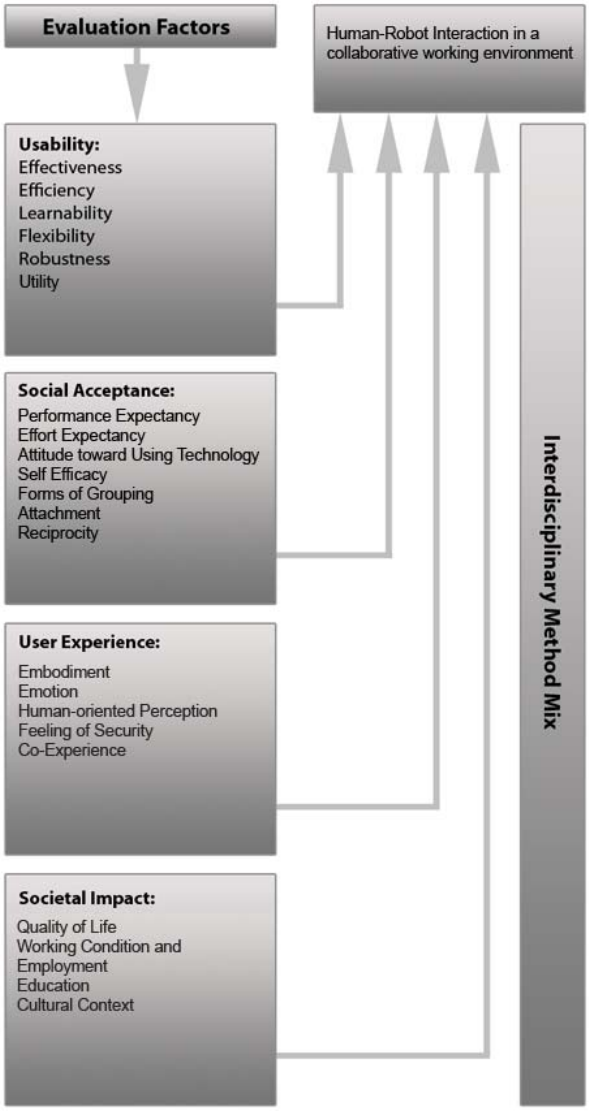


Figure 1: The Evaluation Framework

3.1 Usability as Evaluation Factor

The term usability refers to the ease of using an object. The iso924111:1998 [27] defines usability as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use”. This definition shows that usability is a concept of different indicators, than one single measurable term. Initial research on usability in human-robot interaction was mainly concentrating on indicators like, performance/ effectiveness and efficiency (e.g. [14], [40]). However, we propose that further aspects should be taken into account when assessing the usability of a humanoid robot.

3.1.1 Indicators for Usability

Effectiveness: The iso924111:1998 [27] defines effectiveness as "the accuracy and completeness with which users achieve specified tasks". Thus effectiveness describes how well a human-robot team accomplishes some task. This normally refers to the degree to which errors are avoided and tasks are carried out successfully solved, measured by e.g. "success rate" or "task completion rate".

Efficiency: In the iso924111:1998 [27] efficiency is defined as "the resources expended in relation to the accuracy and completeness with which users achieve goals". So efficiency is the rate or speed at which a robot can accurately and successfully assist humans.

Learnability: is one indicator of usability derived from software engineering. The concept of learnability is self-explanatory: how easy can a system be learned by novice users. This seems to be a key indicator for usability in human-robot interaction as robots are a technology people have almost no pre-experience with. Learnability incorporates several principles like familiarity, consistency, generalizability, predictability, and simplicity.

Flexibility: As humanoid robots are in general task independently designed (as they should be able to carry out a variety of tasks in unstructured environments and adapt to situations), flexibility seems to be another core-indicator for the usability evaluation of humanoid robots in collaborative working environments. Flexibility describes the number of possible ways how the user can communicate with the system.

Robustness: Novice users will produce errors when collaborating with humanoid robots, thus an efficient human-robot interaction has to allow the user to correct its faults on his/ her own. Furthermore the robotic system itself should be error preventing, by means of being responsive and stable. Robustness is thus the level of support provided to the user to enable a successful achievement of tasks and goals.

Utility: As usability relates to the question of effectiveness and efficiency, like how well an interface supports the user to reach a certain goal or to perform a certain task, utility refers to how an interface can be used to reach a certain goal or to perform a certain task. The more tasks the interface is designed to perform, the more utility it has. Therefore utility and usability are related, but not interchangeable. Regarding humanoid robots utility is an essential factor, as a novice user has little knowledge about the utility of this type of robot as they are not designed for a specific task.

3.2 Social Acceptance as Evaluation Factor

Acceptance is an important issue to be evaluated in human-centred HRI. There is a need to find out the reasons why people accept robots in order to avoid rejection in a long term. Dillion [11] defines user acceptance as "the demonstrable willingness within a user group to employ technology for the tasks it is designed to support". However, the acceptance of autonomous acting robots cannot be defined that easily. In Western cultures a general retention of autonomous robots is present (e.g. [26], [30]) and furthermore novice users have difficulties in interpreting for which tasks a robot is designed. Thus, for socially situated robots [14], which can perceive and react on a social environment a different view on the term acceptance is necessary. Social acceptance within the USUS evaluation framework is defined as "an individual's willingness based on interaction experiences to integrate a robot into an everyday social environment". Several acceptance models exist which propose a theoretical framework for investigating technology acceptance, an excellent overview can be found in [45].

3.2.1 Indicators for Evaluating Social Acceptance

The indicators for the factor social acceptance in the USUS framework are derived from the UTAUT (Unified Theory of Acceptance and Use of Technology) model [46] (indicators 1 to 4) and from the theory of "object-centred sociality" [33] (indicators 5 to 8). The indicators are defined in accordance with the theory of object-centered sociality to understand the important aspect of how humans can be socially influenced in their working routines by a robot.

Performance Expectancy: In accordance to the UTAUT model Performance Expectancy is the strongest predictor of usage intention (it is significant at all points of measurement during the development of the model, both in voluntary and mandatory settings). "Performance expectancy is defined as the degree to which an individual believes that using the system will help him or her to attain gains in job performance." [46]..

Effort Expectancy: indicates to which extent the user perceives a system will be easy to use. Thus it includes believes of the degree of effort, difficulties, and understanding in usage, but also how complex users imagine the system to be. "Effort expectancy is defined as the degree of ease associated with the use of the system." [46].

Attitude toward Using Technology: In the UTAUT model the attitude toward using technology is defined as "an individual's overall affective reaction to using a system" [46]. In this evaluation framework attitude toward using technology is seen as the sum of all positive or negative feelings and attitudes about solving working tasks supported by a humanoid robot.

Self Efficacy: relates to a person's perception of their ability to reach a goal. This indicator is not included as a direct determinate in the UTAUT [46] but it is estimated to be a relevant factor for human-robot interaction. Perceived self efficacy can be defined as "people's beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives. Self Efficacy beliefs determine how people feel, think, motivate themselves and behave. Such beliefs produce these diverse effects through four major processes. They include cognitive, motivational, affective and selection processes" [3].

Forms of Grouping: Group practices are a core element of human behaviour. Grouping describes that humans who share certain characteristics, try to collect in a group. They interact preferably with other group members, share a common identity, and accept expectations and obligations of other group members. The question arising is whether humans can also share identity with robots.

Attachment: The term attachment was originally used to explain the bond that develops between a human infant and its caregiver [6]. In the last decades, the concept of emotional attachment has been used in a number of ways, also in relation to HRI (e.g. [29], [28]). Attachment may be defined as an affection-tie that one person forms between him/ herself and another person or object - a tie that binds them together in space and endures over time. According to Norman [38], emotional attachment can be seen as the sum of cumulated emotional episodes of users' experiences with a device in various context areas. These experience episodes can be categorized into three dimensions: a visceral level (first impression), a behavioural level (usage of the device), and a Reflective level (interpretation of the device).

Reciprocity: describes the principle of give-and-take in a relationship, but it can also mean the mutual exchange of performance and counter-performance. It is the positive or negative response of individuals towards the actions of others.

3.3 User Experience as Evaluation Factor

The term user experience is a very multifaceted concept and the research field of Human-Computer Interaction is still searching for a shared understanding of it [35]. A suitable definition for user experience of human-robot interaction can be adapted from Alben's general definition of user experience as "aspects of how people use an interactive product: the way it feels like in their hands, how well they understand how it works, how they feel about it while they're using it, how well it serves their purposes, and how well it fits into the entire context in which they are using it" [1].

Thus, users' experiences are related to a system and are embedded in a specific situation. Interaction goals, intra-psychological dispositions, the environment, involved people and the product itself have a significant impact on user experience [20].

3.3.1 Indicators for Evaluating User Experience

There is an increased interest in HRI establishing a positive experience for the interaction with a robot. In a working environment a positive user experience is desired since working tasks will be carried out more efficiently. In the following some factors of user experience in HRI are introduced, which are mainly derived from the framework of [15] to classify socially interactive robots.

Embodiment: describes the relationship between a system and its environment and can be measured by investigating the different perturbatory channels like morphology, which has impact on social expectations [15]. It is assumed that a humanoid form will ease the human-robot interaction because the rules for human social interaction will be invoked and thus a humanoid robot will provide a more intuitive interface, although this premise is largely untested. This is an assumption which is often found in literature (e.g. [7], [8]). Other researchers assume that robots are perceived as machines and that humanoid features therefore would generate unrealistic expectations or even fear [15].

Emotion: The indicator emotion implies that people tend to interact with computers (and robots) socially [39]. As emotion is an essential part in social interaction it has to be incorporated in the assessment and design of robots. Hassenzahl [19] structured aspects of user experience in product, goals related to the product, psychological status of a user, environment and people related to the experience. He stressed the importance of emotion in user experience by introducing relevant emotional episodes that are aroused during the interaction with a product. Users may experience satisfaction when a product fulfils the users' expectations. The emotion joy is felt by exceeding the user's expectations. Furthermore pride, surprise and attraction play a major role in experiencing a product.

Human-Oriented Perception: tries to simulate human perception. Social robots should be capable of tracking human features (e.g. face), interpreting human speech and recognize facial expressions etc. Robots should have different types of perception, meaning passive sensing and spoken language recognition. They have to be able to track people in different environment and lighting conditions. Additionally, the system should be able to recognize speech in two steps: speech processing and graph search. Furthermore vision based gesture recognition and facial perception (face detection and recognition) skills are necessary to guarantee Human-Oriented Perception. The system should not only be able to recognize facial display but also should have capabilities for communicating facial expression by means of image motion techniques, anatomical models, and principal component analysis.

Feeling of Security: As soon as humans should collaborate together with robots in the same environment, safety and security issues arise [10]. In addition to studies on how to eliminate the risk of hazards in human-robot collaboration [34], it is important to investigate how to design human-robot interaction in a way that humans experience them as safe. For example, [10] discovered that people prefer to be approached by a robot on the right hand side.

Co-Experience with Robots: Co-experience describes experiences with objects regarding how individuals develop their personal experience based on social interaction with others. People define "situations through an interpretive process in which they take into account the non-symbolic gestures and interpretations of others" [5]. Severinson-Eklund et al. e.g. [41] observed collaborative aspects of interaction with robots focusing on the personality of the robot, the communication paradigm between the user and the robot, and how a robot can mediate within a group of people.

3.4 Societal Impact as Evaluation Factor

The maturing of technology has been seen ever since as process that influences and changes society (e.g. consequences of the industrialization). Turn [44] conducted one of the first studies on the societal impact of computing technology. Turn pointed out that the purpose of societal impact studies should not only be to analyze the actual state of society, but to identify potential problems of future society, and to recommend corrective actions.

Societal impact can be defined as every effect of an activity on the social life of a community in general and more specific for the proposed framework: "Societal impact describes all effects the introduction of robotic agents consequences for the social life of a specific community (taking into account cultural differences) in terms of quality of life, working conditions and employment, and education."

Theoretical assumptions on how the future society could look like and be influenced by robotic agents can above all be found in the cyborg and post-humanism literature (e.g. [2], [17]).

3.4.1 Indicators for Societal Impact

One of the main challenges when evaluating HRI is to predict the societal impact of robots. Already in 1981 the office of Technology Assessment conducted an exploratory workshop with the aim "to examine the state of robotics technology and possible public policy issues of interest". Participants of this workshop identified four areas of "social issues" relevant for future societies in terms of the integration of robotic technology: (1) productivity and capital formation, (2) labor, (3) education and training, (4) international impact. Similar relevant impact factors could be derived from post-humanism and cyborg literature

Quality of Life, Health and Security: According to Gray [17] quality of human life is determined by several types of freedom, like free choice of gender orientations, or the freedom of travel. Furthermore, Gray argues that also stable human relationships, family constellations, and mutual reliance have important impact on human life quality and that the very nature of our relationships with each other will change through the integration of artificial intelligence into our environments.

Also the health system will be influenced by these developments, as high tech medicine will allow new therapy possibilities, which go hand in hand with the possibility of living longer. Security aspects could be concerned by the integration of intelligent robotic technology into everyday life, like electronic privacy, the freedom of consciousness, and the freedom of information. Already examples exist giving hints on these possible tendencies in future societies (e.g. [13], [16]). These researches efforts can on the on hand support the future health system and thus improve the quality of life, but on the other hand it could harm the natural relationships.

Working Conditions and Employment: Working conditions and employment includes all aspects affecting how people carry out their job and how employers take care of their employees, including things like working contracts, wages, working times, and work organization. Working conditions have always been affected by technology developments, as they can be used to increase the efficiency and productivity. This in turn may lead to an increasing degree of replacement of e.g. assembly-line workers by robots, as robots can complete some physical tasks much quicker and more precisely than humans, e.g. harvesting [18]. Forlizzi et al. e.g. think of closing the gap of lacking service personal in hospitals by the introduction of care robots, for situations when no physical presence of a doctor or a nurse is necessary [16].

Education: New software, new sciences and new disciplines require new types of education. Lifelong learning is necessary to manage duties and responsibilities. To avoid the fear of being displaced by a robot it might be necessary to give educational advertising. In times of increasing utilization of robots the aspect of education should not be disregarded. Considering the situation now, after the launch of computers, every child in western society is taught how to use computers and how to use certain programs. But without this education most people would not be prepared sufficiently for the labour market. So it might be necessary to be prepared for utilization of robots, in the physical manner, but potentially in a psychological manner too.

Cultural Context: Culture embraces the whole range of practices, customs and representations of a society. In their rituals, stories and images, societies identify what they perceive as good and evil, proper, and racially different [2]. However, culture does not exist in the abstract. On the contrary, it is in the broadest sense of the term textual, inscribed in the paintings, operas, sculptures, furnishings, fashions, bus tickets and shopping lists which are the currency of both aesthetic and everyday exchange [2]. Thus the socio-cultural environment plays a decisive role. Japanese or South Koreans interact with robots in a quite different and more enthusiastic manner than in Europe, where people are more sceptical. This is due to the fact that in Japan e.g. automatons have a long tradition in religious ceremonials. Furthermore, in special Japanese religions a soul to things and machines is granted. Last but not least the positive presentation of robots in Japanese literature leads to a high acceptance too. There are also great differences in Japan and Europe regarding robots in the working area. Japanese employees trust in the corporation's decision and establish a long and quasi familiar relationship to their corporation. In contrast, in Europe due to short employment contracts and numerous structural changes over the last years robots are often perceived as a rationalization instrument [26].

4 THE METHODOLOGICAL FRAMEWORK

Different methods are available to assess interactive systems in general and human-robot collaboration in specific. Depending on the data that is required, the resources that are available (i.e., time and money) and the design phase, a decision can be made which methods to choose [12]. One of the basic principles of a user-centred evaluation approach is that potential users evaluate whether the system is usable, acceptable etc. or not. Moreover, the method selected has to be adapted to the context, the tasks and the system which will be evaluated. Formative evaluation approaches are the main interest of this framework, as summative evaluations in real workplace environments with humanoids are hardly possible.

To investigate and evaluate the above defined indicators for usability, social acceptance, user experience and societal impact, a combination of different methods is needed. Qualitative research is combined with quantitative measures for the evaluation approach. The proposed methods, presented in a matrix-diagram visualizing which method is suitable for which indicator, are presented later on in more detail.

Table 1: The Methodological Mix

	Methods	Expert Eval	User Studies	Questionnaires	Physio. Measures	Focus Groups	Interviews
Research Objectives							
Usability							
	Effectiveness	X	X				
	Efficiency	X	X				
	Learnability	X	X				
	Flexibility	X	X				
	Robustness	X	X				
	Utility			X			X
Social Acceptance							
	Performance Expectancy			X		X	
	Effort Expectancy			X		X	
	Attitude toward Using Technology			X			
	Self Efficacy			X		X	
	Forms of Grouping			X		X	
	Attachment			X		X	
	Reciprocity			X			
User Experience							
	Embodiment			X		X	
	Emotion			X	X	X	
	Human-Oriented Perception			X			
	Feeling of Security			X	X	X	
	Co-Experience			X		X	
Societal Impact							
	Quality of Life			X		X	X
	Working Conditions			X		X	X
	Education			X		X	X
	Cultural Context			X		X	X

4.1 Expert Evaluation

In traditional HCI research expert evaluations are used to asses a system in terms of its usability and detect as many usability problems as possible in a way that is less cost- and effort-intensive than user testing.

Heuristic Evaluation: A heuristic evaluation is intended to find and describe usability problems of a system on the basis of fundamental principles, so called heuristics [37]. Heuristics are describing essential attributes that a system should feature to ensure that the user is able to perform a task within a specified context in an effective, efficient, and satisfying way. Such a heuristic evaluation is usually performed by a small team of interface experts inspecting the system and comparing to what extent the principles have been adopted. All experts then have to rank all problems according to their severity. Thus, the result of a heuristic evaluation is a complete list of all detected usability problems ranked according to their severity.

Cognitive Walkthrough: A cognitive walkthrough [47] is conducted by at least two usability experts assessing the usability of a system based on predefined task structures. The expert evaluators try to imagine how a typical (potential) user would solve a task with the assumption of minimizing the cognitive load. Thus, the cognitive walkthrough is above all used to evaluate the usability of a system in terms of its learnability and how intuitive it can be used. During each task analysis the experts ask themselves a set of questions for each subtask.

4.2 User Studies

Laboratory-based: User studies are used to provide empirical evidence to answer a concrete research question or hypothesis. Such user-involving evaluations are based on tasks, which subjects conduct, while their behaviour is observed and measured by a researcher. Classical metrics measured during user studies are task completion rate or error rate to measure the effectiveness of the tested system, and task duration to measure the efficiency. To better understand usability problems of the subjects occurring during the solving of task the "think aloud" method is often applied in user studies. "Think aloud" means that subjects are asked to say whatever they are looking at, thinking of, doing, and feeling, as they conduct a task. This enables observers to see first-hand protocol usability problems. User studies are normally audio and video taped so that researchers can go back and refer to what subjects did, and how they reacted. User studies provide a good opportunity to be combined with other methods like questionnaires or qualitative interviews.

Field-based: Field based user studies have the focus of interest in testing the usage of a system in a realistic usage context. This kind of user studies can, but does not necessarily have to be task-based. In general the procedure is similar to user studies conducted in the laboratory, whereas the observation is mostly passive and unstructured. However, as field trials have to take into account more disturbing factors (e.g background noise or light conditions), the interpretation and analysis of the data, is more difficult.

Wizard of OZ Technique: User studies can be conducted with fully autonomous systems, or they can be based on the so-called "Wizard of OZ Technique", short "WOZ". [31]. To allow user testing in very early stages of the prototype development, which cannot be fully implemented at that stage, a human "Wizard" enacts (or simulates) the system features in interaction. This approach integrates several advantages for user studies in HRI, as e.g. safety and security issues can be controlled during the testing, and relevant social cues can be simulated. However, also disadvantages are incorporated when evaluating user experience and social acceptance: Is the perception of the robot measured, or the perception of a human wizard...?

4.3 Standardized Questionnaires

A questionnaire is a research instrument that consists of a series of questions with the purpose of gathering statistically analyzable data from the participants. Standardized questionnaires are based on closed answers, where participants have only to choose one of the pre-defined answers making it easier for them to complete the questionnaire. A good overview on existing questionnaires for HRI research can be found in [2].

4.3 Physiological Measurements

Physiological measurements can give valuable additional input to e.g. questionnaires and focus groups and can detect information participants for instance do not want to state. Several methods can be used to measure the emotional state of a subject [36]: Methods like this can identify “emotions” at exactly the time they happen during a user study and can be combined with e.g. reflective questionnaire data on the emotional state. This can support investigation of user experience factors in terms of “heteronom and autonom identification”.

4.4 Focus Groups

Focus groups allow the researcher to explore participants’ attitudes, beliefs, and desires in great depth and give insights how they experience a system. Focus groups are structured discussions about specific topics and moderated by a trained leader. The focus of the discussion is also triggered by the participants' selection which is based on common characteristics, as opposed to differences among them. It is important to note that a focus group only gathers qualitative data, which can be used as input to further develop other research instruments which gather quantitative and generalizable results.

4.5 In-depth Interviews

In-depth interviews are a qualitative research technique, which allows “person-to-person” discussion on a specific topic. It aims to get increased insight participant's ideas, attitudes, and feelings on the discussed issues. In-depth interviews can be combined with user studies to discuss with the participants how they experienced the interaction with the robotic system.

Expert Interviews: This specific type of in depth interview is conducted in the same way as qualitative interviews in general. However, they are conducted with people who are considered experts in a particular subject, e.g. humanoid robot development, robots in the workplace, robots and ethics etc. The aim is not to get to know their attitude or feelings on a topic, but that they share their knowledge.

Delphi study: is a special form of how to conduct expert interviews. The goal is to find a solution for a complex problem statement. Delphi studies consist of several rounds of interviews and discussions with several experts. After each round researchers report to experts the results of the previous discussion round as group opinion about the problem statement. This process happens as long as a common solution for the problem statement is found.

5 SUMMARY AND OUTLOOK

The goal of the framework is a multi-level evaluation model covering a multitude of factors: Usability, social acceptance, user experience, and societal impact. These four factors are called the USUS factors. Main goal of the evaluation framework is to guide research in answering the question on how people experience robots as a support for collaborative work and accept them as part of society. The framework operationalize the USUS factors with indicators and describes methodological possibilities to investigate them - and beyond that, helps to The evaluation framework is intended to guide current activities within the Robot@CWE research project, but it can help other researchers on a more general level to understand what kind of methods can be helpful to investigate the USUS factors in human-robot interaction. To assess the validity of the proposed framework currently more than twenty evaluations, at seven different sites involving various types of humanoid robots, are conducted to investigate the relationship between the various factors (first results can be found in ([49][50][51]). Further results will be available in 2009, showing advantages and limitations of the proposed framework to evaluate human-robot collaboration with humanoid robots, but we expect this framework to be a valuable help for researchers to investigate the USUS factors in HRI.

ACKKNOWLEDGEMENTS

This framework was developed in the framework of the ROBOT@CWE project (funded by FP6-2005- IST-5). The authors like to thank all partners from the project consortium.

REFERENCES

[1] L. Alben, ‘Quality of experience: defining the criteria for effective interaction design’, interactions, 3(3), 11–15, (1996).

[2] N. Badmington, Posthumanism, Readers in cultural criticism, Palgrave, Houndmills, Basingstoke, Hampshire , New York, 2000.

[3] A. Bandura, Self-efficacy: The exercise of control, New York, NY, Freeman, 1997.

[4] C. Bartneck, E. Croft, and D. Kulic, ‘Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots’, International Journal of Social Robotics, (2009).

[5] K. Battarbee, ‘Defining co-experience’, in DPPI ’03: Proceedings of the 2003 international conference on Designing pleasurable products and interfaces, pp. 109–113, New York, NY, USA, (2003). ACM.

[6] J. Bowlby, ‘The nature of the child’s tie to his mother’, *International Journal of Psychoanalysis*, **39**, 350–373, (1958).

[7] C. Breazeal and B. Scassellati, ‘How to build robots that make friends and influence people’, in *IROS ’99: IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 2, pp. 858–863, (1999).

[8] R. Brooks, ‘Humanoid robots’, *Commun. ACM*, **45**(3), 33–38, (2002).

[9] K. Dautenhahn, ‘The art of designing socially intelligent agents : Science, fiction, and the human in the loop’, *Applied Artificial Intelligence*, **12**(7), 573 – 617, (1998).

[10] K. Dautenhahn, M. L. Walters, S. Woods, K. L. Koay, E.A. Nehaniv, C. L. Sisbot, R. Alami, and T. Simeon, ‘How may I serve you?: a robot companion approaching a seated person in a helping context’, in *HRI ’06: Conference on Human-Robot Interaction*, pp. 172–179, Salt Lake City, Utah, USA, (3 2006). ACM.

[11] A. Dillon, ‘User acceptance of information technology’, in *Encyclopedia of Human Factors and Ergonomics*, ed., W. Karwowski, Taylor and Francis, London, (1 2001).

[12] A. Dix, G. D. Abowd, and J. E. Finlay, *Human-Computer Interaction (3rd Edition)*, Prentice Hall, 2004.

[13] R. Dobson, 'Meet Rudy, the world's first "robodoc"', *BMJ*, **329**(7464), 474, (2004). 1 University of Leipzig, Germany, email: somename@informatik.unileipzig. de

[14] J. L. Drury, J. Scholtz, and H. A. Yanco, 'Awareness in human-robot interactions', in *IEEE Conference on Systems, Man and Cybernetics*, pp. 912– 918, Washington, (10 2003).

[15] T. Fong, I. Nourbakhsh, and K. Dautenhahn, 'A survey of socially interactive robots', *Robotics and Autonomous Systems*, **42**, 143–166, (3 2003).

[16] J. Forlizzi, 'Robotic products to assist the aging population', *interactions*, **12**(2), 16–18, (2005).

[17] C. H. Gray, Cyborg citizen: Politics in the posthuman age, Routledge, New York, 2001.

[18] S. Green, X. Q. Billinghurst, M.and Chen, and G. Chase, 'Human-robot collaboration: A literature review and augmented reality approach in design', *International Journal of Advanced Robotic Systems*, **5**(1), 1– 18, (2008).

[19] M. Hassenzahl, 'The effect of perceived hedonic quality on product appealingness', *Int. J. Hum. Comput. Interaction*, **13**(4), 481–499, (2001).

[20] M. Hassenzahl, 'The thing and i: understanding the relationship between user and product', in *Funology. From Usability to Enjoyment*, eds., M. Blythe, C. Overbeeke, A. F. Monk, and P. C. Wright, 31–42, Kluwer, Dordrecht, (2003).

[21] M. Heerink, and B. Kröse, B. Wielinga, and V. Evers, 'Studying the acceptance of a robotic agent by elderly users', *International Journal of Assistive Robotics and Mechatronics*, **7**(3), 25–35, (9 2006).

[22] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, 'The influence of a robot's social abilities on acceptance by elderly users', in *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pp. 521–526, (2006).

[23] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, 'The influence of social presence on enjoyment and intention to use of a robot and screen agent by elderly users', in *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, pp. 695–700, (2008).

[24] M. Heerink, B. Kröse, B. Wielinga, and V. Evers, 'Enjoyment intention to use and actual use of a conversational robot by elderly people', in *HRI '08: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pp. 113–120, New York, NY, USA, (2008). ACM.

[25] P. J. Hinds, T. L. Roberts, and H. Jones, 'Whose job is it anyway? a study of human-robot interaction in a collaborative task', *Human-Computer Interaction*, **19**(1-2), 151–181, (2004).

[26] T. N. Hornyak, *Loving the machine: the art and science of Japanese robots*, Kodansha international, Tokyo, New York, London, 2006.

[27] ISO 9241-11, Ergonomic requirements for office work with visual display terminals - Part 11: Guidance on usability, International Organization for Standardization, 1998.

[28] F. Kaplan, 'Free creatures: The role of uselessness in the design of artificial pets', in *Proceedings of the 1st Edutainment Robotics Workshop*, (September 2000).

[29] F. Kaplan, 'Artificial attachment: Will a robot ever pass the ainsworth's strange situation test?', in *Proceedings of Humanoids 2001: IEEE-RAS International Conference on Humanoid Robots*, pp. 125 – 132, (2001).

[30] F. Kaplan, 'Who is afraid of the humanoid? investigating cultural differences in the acceptance of robots', *International journal of humanoid robotics*, **1**(3), 465–480, (2004).

[31] J. F. Kelley, 'An iterative design methodology for user-friendly natural language office information applications', *ACM Trans. Inf. Syst.*, **2**(1), 26–41, (1984).

[32] S. Kiesler and P.J. Hinds, 'Introduction to this special section on human-robot interaction', *Human-Computer Interaction, Spec. Issue on Human-Robot Interaction*, (1-2), (2005).

[33] K. Knorr-Cetina, 'Sociality with objects: Social relations in postsocial knowledge societies', *Theory Culture Society*, **14**(4), 1–30, (November 1997).

[34] D. Kulic and E. Croft, 'Strategies for safety in human robot interaction', in *ICAR'03: IEEE International Conference on Advanced Robotics*, pp. 644–649, Coimbra, Portugal, (2003).

[35] E. Law, V. Roto, A. Vermeeren, J. Kort, and M. Hassenzahl, 'Towards a shared definition of user experience', in *CHI '08: CHI '08 extended abstracts on Human factors in computing systems*, pp. 2395–2398, New York, NY, USA, (2008). ACM.

[36] M. Minge, Methoden zur Erhebung emotionaler Aspekte bei der Interaktion mit technischen Systemen, Ph.D. dissertation, Freie Universitaet Berlin, Berlin, 2005.

[37] J. Nielsen, 'Finding usability problems through heuristic evaluation', in *Proceedings of the ACM CHI 92 Human Factors in Computing Systems Conference*, pp. 373–380. ACM Press, (1992).

[38] D. A. Norman, Emotional Design: Why We Love (Or Hate) Everyday Things, Basic Books, New York, USA, 2004

[39] B. Reeves and C. Nass, The Media Equation: How People Treat Computers, Televisions, and New Media Like Real People and Places, Cambridge University Press, New York, 1996.

[40] J. Scholtz, 'Evaluation methods for human-system performance of intelligent systems', in *PerMIS'02: Performance Metrics for Intelligent Systems Workshop*, Gaithersburg, MD, USA, (8 2002).

[41] K. Severinson-Eklundh, Anders Green, and Helge Hüttenrauch, 'Social and collaborative aspects of interaction with a service robot', *Robotics and Autonomous Systems*, **42**(3-4), 223–234, (2003).

[42] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, 'Common metrics for human-robot interaction', in *HRI '06: Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pp. 33–40, New York, NY, USA, (2006). ACM.

[43] S. Thrun, 'Toward a framework for human-robot interaction', *Human-Computer Interaction*, **19**(1), 9–24, (2004).

[44] R. Turn, 'Courses on societal impacts of computers', *SIGCAS Comput. Soc.*, **13, 14**(4, 1-3), 14–16, (1984).

[45] V. Venkatesh and F. D. Davis, 'A theoretical extension of the technology acceptance model: Four longitudinal field studies', *Manage. Sci.*, **46**(2), 186–204, (2000).

[46] V. Venkatesh, M. G. Morris, G. B. Davi, and F. D. Davis, 'User acceptance of information technology: Toward a unified view', *MIS Quarterly*, **27**(3), (2003).

[47] C. Wharton, J. Bradford, R.Jeffries, and M. Franzke, 'Applying cognitive walkthroughs to more complex user interfaces: experiences, issues, and recommendations', in *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 381–388, New York, NY, USA, (1992). ACM.

[48] H.A. Yanco and J. Drury, 'Classifying human-robot interaction: an updated taxonomy', *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, **3**, 2841–2846, (Oct. 2004).

[49] A. Weiss, R. Bernhaupt, M. Tscheligi, D. Wollherr, K. Kühnlenz, and M. Buss. Methodological variation for acceptance evaluation of human-robot interaction in public places, in *IEEE RO-MAN 2008: Proceedings of the International Symposium on Robot and Human Interactive Communication,* Munich, Germany, (2008).

[50] A. Weiss, R. Bernhaupt, M. Tscheligi, and E. Yoshida. 'Addressing user experience and societal impact in a user study with a humanoid robot', in *AISB2009: Proceedings of the Symposium on New Frontiers in Human-Robot Interaction (accepted for publication)*, (2009).

[51] A. Weiss, D. Wurhofer, M. Lankes, and M. Tscheligi. 'Autonomous vs. tele-operated: How people perceive human-robot collaboration with HRP-2', *poster presentation at HRI 2009:ACM/IEEE International Conference on Human-Robot Interaction*, (2009).