FP6-IST-002020

COGNIRON

The Cognitive Robot Companion

Integrated Project

Information Society Technologies Priority

D3.2005
Social Behaviour and Embodied Interaction

Due date of deliverable: 31/12/2005
Actual submission date: 25/01/2006

Start date of project: January 1st, 2004 **Duration**: 48 months

Contributing partners: UH, KTH, LAAS, UniKarl, UvA

Revision: Final
Dissemination level: PU

Executive Summary

During 2004 and 2005 work in RA3 has developed according to the work specified in the second implementation plan. In line with the recommendations at the last review, there are increased efforts in the area of in-depth analysis of HRI studies relevant to the KEs that inform other RAs. During 2004 and 2005 HRI trials were carried out at UH and KTH investigating human-robot social spaces, robot behaviour styles and posture and positioning in HRI. Within WP3.1 scenarios involving negotiation of space, robot assistance provided in a helping task, and robot-to-human approach directions were investigated in collaboration with WP3.3 on robot motion planning and relevant specifically to KE2 (Curious Robot scenario). In WP3.2, examples of spatial management during “Follow”, “Show”, and “Validate” joint-task observations directly relevant to KE1 (Home Tour) were collected and analyzed. During the first months of 2005, WP3.1 and WP3.2 concentrated on in-depth analysis of empirical data and dissemination of results to other partners and RAs as well as publication of results in line with the second implementation plan. Within WP3.3 LAAS developed a navigation planner in simulation, integrating results from WP3.1 that will in 2006 be implemented and tested on a robotic platform. The new WP3.4 that started in June 2005, as a response to reviewers’ comments, now provides a strong link to RA2 in order to integrate the development of recognition algorithms for human activities and human gestures to empirical data on ‘naturally’ occurring (unscripted) human behaviour in HRI trials. UH and UniKarl collaborate in WP3.4 to investigate gestures and body movements in a human-robot teaching scenario relevant to KE3 (Skill Learning). Joint work between UH and UvA considers social spaces in a following and guidance scenario relevant to KE1. A user study has been designed for annotating video data with human comfort levels. KTH’s work in WP3.4 is closely associated with KE1 and along with the work in WP3.2 focuses on the human-robot dialogue studied in RA1 in collaboration with University of Bielefeld (UniBi). The focus of this work during the second period has consisted of the development of an annotated multi-modal corpus of interaction sequences in an initial work on taxonomy of gestures and body movements, based on empirical data. During this implementation period, dissemination of outcomes from RA3 have included two technical reports and twelve papers being presented or accepted for publication at various international conferences including IEEE CIRA 05, IEEE Ro-man 05, IROS 05, Humanoids 05, AISB’05 Robot Companion Symposium and HRI’06. Of these papers, three are joint publications; UH/IPA/LAAS at Ro-man 05, LAAS/UH at Humanoids 05 and UH/LAAS for one of the papers accepted for HRI’06. There has also been one journal submission with several more planned for 2006.

Table of Contents

Role of Social Behaviour and Embodied Interaction in COGNIRON

In the context of HRI social behaviour and embodied interaction is an important area of research that involves numerous issues of e.g. verbal, non-verbal and affective interaction. During the first two project years a few specific research topics have been identified and investigated in RA3 that were considered most relevant for meeting the project's objectives of studying a cognitive robot companion. Research Area (RA) 3 is concerned with social spaces, gestures, postures and body movements occurring in human-robot interactions and their role in research on robot motion planning, navigation and recognition of human-activities. This work is split into four Work Packages (WPs) which aim at providing scientific insights based on experimental data on socially acceptable, primarily non-verbal behaviour as relevant to the Key Experiments (KEs). Verbal communication (dialogue) and affective factors (user comfort) are considered as far as they are relevant for the specific scenarios in the KEs. Different but complementary methods and methodologies are used to investigate these issues; experimental data is derived from HRI user studies as well as simulation and robotics testbeds. The work performed contributes to the development of frameworks and models of interaction styles for Human-Robot Interaction (HRI) that a) inform the implementation and testing of the KEs within COGNIRON in order to achieve the project's objectives, and b) via scientific dissemination contribute to knowledge in the wider scientific HRI and robotics community.

Relation to the Key Experiments

The results from the work in RA3 are intended to support KEs 1, 2 and 3 primarily with regard to the following COGNIRON Functions as described in the Key Experiment Specification Documents [COGNIRON, 2005a-c].

- CF-SOC: Socially acceptable interaction with regard to spaces.
- CF-DLG: Multimodal dialogue including robust speech understanding for natural communication.
- CF-NHP: Navigation in human presence
- CF-ACT: Detection and interpretation of human activities and postures
- CF-PTA: Person tracking and detection of attention
- CF-MHP: Manipulation in human presence

RA3 investigates issues that are relevant across the different Key Experiments. Individual experiments within the individual WP's typically focus on scenarios relevant to a specific KE. WP3.1 and WP3.2 primarily target KE2 ("Curious robot") and KE1 ("Home tour") respectively. WP3.3 is mostly relevant for KE2 but also wider issues of motion planning in the presence of humans which can be found across the KEs. WP3.4 targets primarily KE1 and KE3 ("Skill learning") due to the nature of the experiments that are linked to other RAs.

RA3: Social Behaviour and Embodied Interaction

This deliverable covers the work carried out to achieve the aims of Research Area 3 in 2005 for the COGNIRON project. It provides a summary of the research activities carried out to fulfil the objectives outlined in the second implementation plan for WPs 3.1, 3.2, 3.3, and 3.4. It also provides a documentary record of activities, outcomes and dissemination within the project and in relation to peer reviewed scientific papers and conference activities. Future research for the next planning period will be mentioned briefly.

The main changes to the WPs for RA3 second project phase have resulted from the reduced involvement of VUB and IPA in RA3. As outlined in the revised WPs for RA3 for Months 12-18, efforts have also been taken to strengthen RA3 by increasing the effort invested by the existing partners and also by involving other partners (UvA and UniKarl) previously not involved in RA3, and also by adding a new workpackage (WP3.4) that specifically links research in RA2 (Detecting and Understanding of Human Activity) and RA3 in response to the feedback received from the reviewers at the first review. The WPs for the second implementation plan emphasised the scientific integration within RA3 with other research activities and the relationships to the Key Experiments. Both WP3.1 and WP3.2 involve extensive user studies which address hard problems in social behaviour and embodied interaction in experimental robotic setups involving mobile robots and human subjects. The main themes were to further develop from previous work and to move on from first exploratory studies to more focussed research questions, in-depth analysis of results and an extension of the scope of the scenarios. Also, it was a desirable aim to add more autonomy to the robots used in these trials, complementing studies in the first implementation phase which were primarily Wizard of OZ (WoZ) controlled. WP 3.3 focuses on the development and implementation of algorithms and models for sharing spaces when a robot operates in close proximity to humans. In 2005 a navigation planner was finalised taking into consideration results of user studies performed in WP3.1. This planner is now being implemented on a robotics platform so that results can in future be tested and verified in HRI and robotics experiments.

A completely new WP 3.4 has been introduced from June 2005 which explores requirements for contextual interpretation of body postures and human activities. This new WP has already demonstrated the potential to establish scientific integration between research in RA2 and RA3, as well as HRI studies carried out at UH and KTH. It has also started to provide input for research on recognition and tracking of human activities required for RA2 and also has established links to RA1 and RA7. WP3.4 provided an opportunity to scientifically ground research on recognition and tracking of human activities in HRI studies investigating ‘naturally occurring’ (not pre-defined or pre-scripted) gestures, postures and body movements. Moreover, while a comprehensive investigation of perception, expression and modelling of emotions and affect goes beyond the scope of COGNIRON, some affective aspects were included as far as they are relevant to the KE scenarios studied.

The current overall structure of RA3 in the 2nd Implementation Plan (M13-M24) has been reflected in the four WPs which help define the research work areas:

- WP3.1 Personal spaces and social rules in human-robot interaction
- WP3.2 Posture and Positioning

- WP3.3 Models and algorithms for motion in presence and in the vicinity of humans
- WP3.4 Requirements for contextual interpretation of body postures and human activities (WP started in June 2005)

The current progress, achievements, outcomes and future work for each of the four WPs is given in the following sections. While the WPs are overlapping in terms of research issues studied, the WP structure has been maintained for this deliverable in order to allow discussion of specific experiments and results obtained in the individual WPs.

Work Package 3.1: Interaction Styles and Personal Spaces

The University of Hertfordshire (UH) is the lead partner for WP3.1. The main research interest of WP3.1 is in the area of personal spaces and social rules in Human-Robot Interaction. This section will summarise the research activities carried out during 2005 to fulfil the outlined objectives of WP3.1. Substantial time has been spent during 2005 on dissemination activities in relation to WP3.1, including further analysis and publication of results reported in the 2004 deliverable D3.1.1. This is illustrated in the annex of publications provided at the end of this document. As there are a number of publications, a summary of the most important findings will be provided and the reader will be directed to the relevant publication in relation to that particular research if further details are required.

WP3.1: Progress and Achievements in 2005

The period from January to June 2005 was spent primarily completing the analysis and writing up of the results and data obtained from the first two empirical exploratory studies of human-robot interaction in the context of an initial encounter with a robot with a mechanistic appearance performed in 2004. The first was the Children’s Play Scenario (carried out with groups of children) and the second Single Adult Study (with single adults). Both studies were introduced and initial findings presented in the previous COGNIRON Deliverable Report D3.1.1 [COGNIRON, 2005]. Further analysis during the early part of 2005 has allowed the drawing of further interesting results and conclusions from the two studies concerning personal space zones, initial approach distances between robot and humans, the context of the encounters and the human's perception of the robot as a social being. A full discussion of the main results of these observations and analyses can be found in the resultant published peer reviewed papers. Below, a short commentary highlights the main aspects of the research relevant to WP3.1 which is described more completely in the respective papers. In September 2005 a document was produced and circulated to project partners (see the Appendix) which summarises the results of the detailed analyses of the 2004 user studies with respect to social distances. These results specifically informed the Home Tour studies at UniBi. These results and the resulting initial set of 'social rules' for robot behaviour will be refined and extended by results from the new HRI user studies performed in October-December 2005 (analysis of these studies is still ongoing at the time of finalising this deliverable).

A discussion paper by Woods et al. [2005] presented at IEEE CIRA 2005 considered the implications of the psychological phenomena of social facilitation effects for human-robot interaction (HRI) studies. Research studies in HRI have significantly increased over the past few years. Such studies typically investigate e.g. robot appearance and behaviour, and the responses of subjects to robots. However, the possible effects of the experimental context on results from human-robot interaction studies have attracted little attention and an overview of robot trials using children and adults is provided. Observations from video footage are reported, with particular consideration for the influence of the social context and social facilitation effects, including task complexity, evaluation context and type of presence on outcomes of human-robot interaction studies.

Walters et al. [2005a], presented at Humanoids 2005, provides a comparison and discussion of the Human-Robot (HR) and Robot-Human (RH) approach distance data obtained from both the child and adult studies. The child groups showed a dominant response to prefer the 'social zone' distance, comparable to distances people adopt when talking to other humans. From the single adult studies a small majority preferred the 'personal' zone, reserved for talking to friends. However, significant minorities deviate from this pattern. In particular, approximately 40% of the adult sample tested exhibited very close HR approach distances which implies that they saw the robot (which had a mechanistic appearance) simply as a machine, not a social entity.

The results from detailed analysis performed on the video recordings from the child study are presented in te Boekhorst et al. [2005] where two hypothesis were tested: that children are more attentive to a robot if the robot appears to be interested in the children; also if and how the quality and quantity of a child's attentive behaviour varies with the distance to the robot, reflecting the notion of "social spaces". The play scenarios had involved sixteen groups of up to ten children each, in which they had to move closer to a robot over 6 successive rounds. The robot was endowed with a "camera eye" and an arm and hand. The camera was either non-moving ("static") or actively "searching" ("active searching"), giving the impression it was trying to select a child to focus on. The robot's arm and hand could either be fixed in a permanent pointing position ("permanent pointing") or actively rise to point selectively at a particular child when it stopped facing it ("selective pointing"). The main results showed that: 1) The mean frequency of overall attentive behaviour by the children (including attention towards other children) was significantly higher when the robot was not selectively pointing at the children and independent of the state of the camera. 2) "Looking at" was the most frequently scored attentive activity for the children and was mainly targeted to the robot, but not correlated with any of their other attentive activities. 3) There was an interaction effect between the state of the camera and of the pointer: looking at the robot by the children occurred significantly more often when the camera and the arm were consistent in signalling apparent interest (i.e. camera "active searching" and hand "selectively pointing" or camera "static" and hand "permanently pointing"). 4) There was no demonstrable effect of distance to the robot on the overall attentive behaviour of the children.

The practical and methodological lessons learned from carrying out the HRI trials were presented at the 2005 AISB "Robot Companion" Symposium and are discussed fully in Walters et al. [2005b] where methodological, ethical, legal and practical problems involved in conducting human-robot interaction studies are highlighted and critically discussed.

WP3.1: New HRI Trials performed in 2005

During the AISB convention at UH in April 2005, an informal demonstration HRI trial was carried out. Although performed in an uncontrolled reception (party) type environment, the results informed the design of a new HRI study performed in April 2005 involving a robot object fetching scenario. These Video-Live Pilot Studies compared video based vs. real-life HRI trial methods. A paper by Woods et al. [2006] which provides full details from this Video-Live Pilot Study has been accepted for presentation at AMC'06 in March 2006. Subjects participated in both live and video-based HRI trials in which the scenarios were identical, and involved a robot fetching an object using different approach directions. Results of the trials indicated moderate to high levels of agreement for subjects' preferences and opinions for both the live and video based HRI trials. This methodology is in its infancy and should not be seen as a replacement for live trials. However, our results indicate that videotaped HRI trials, in scenarios where the behaviour of the robot is not determined by subjects' reactions (as it was the case for robot to human approach directions) do have potential as a technique for prototyping, testing, developing HRI scenarios, and testing methodologies for use in definitive live trials.

The live trial part of this Pilot Study and the Demonstration Trials had both involved a robot fetching an object that the human had requested using different approach directions. Results from both trials indicated that subjects overwhelmingly disliked a frontal approach, and most preferred to be approached from either the left or right side, with a small overall preference for a right approach by the robot. There was a slight tendency for a few female subjects to prefer a frontal approach, but handedness and occupation were not related to these preferences. The results of the user studies were discussed in the context of path planning for a mobile robot developed by the LAAS team for their work for WP3.3, and especially relevant to the KE2 scenario. The demonstration and pilot trials are described in more detail in Dautenhahn et al. (2006) which has been accepted for presentation at HRI 2006. This paper presents the combined results of both studies investigating how a robot should best approach and place itself relative to a seated human subject. It is jointly authored with the LAAS team, which discusses the results of the user studies in the context of a human aware mobile robot path planning system. These results have now been taken up by the LAAS team for their work for WP3.3 (see section WP3.3) and have now been incorporated in their simulations of a robot navigation system. More details from the perspective of robot motion planning are provided in a joint LAAS/UH publication by Sisbot et al. (2006) which was presented at Humanoids 2005. A navigation planner which has been developed by LAAS takes account of the presence of humans in the robot's environment and in particular humans' safety and comfort. A human-aware motion planner must not only output safe robot paths, but also plan socially acceptable and legible paths. The aim is to build a planner that reasons about a human partner's accessibility, visual field and potential shared motions. The path planner is a part of the development of a human-aware motion planner that will integrate both robot motion and manipulation in a collaborative way with the human.

After these pilot studies, the main set of live HRI Video-Live Approach Direction (VLAD) trials were carried out during October-December 2005. These trials also had the main purpose of expanding and confirming the results obtained from the limited Video Live Pilot study described above, but also including a variety of new and different approach scenarios. We addressed a major criticism of previous trials by participants; namely that the HRI trial environment (lecture or seminar rooms "converted" to a living room) were very "artificial" and "institutional". Although efforts had been made to furnish the experimental areas in such a way as to resemble a home (including flowers, TV etc.), it was still very much apparent that it was a university room. However, in preparation for the longitudinal trials the renting of a residential apartment close to the University, the "Robot House" had been completed by October so that new trials in a much more 'naturalistic' environment could begin. Thus, the VLAD trials, results of which are crucial for the preparation of a new longitudinal study planned for 2006, were run in the new Robot House. Preliminary evidence from the VLAD trials has indicated that the use of the Robot House for the experiments gave the environment a more "home-like" feeling. Live scenarios were conducted by 42 human subjects whereby a UH PeopleBot^{TM} robot autonomously approached and brought an object to the human from a number of different relative approach directions. The 42 subjects of the trial were each tested with two out of four different scenarios. The four scenarios included the subjects: 1) Standing in the middle of a room, (see figure 1), 2) Standing back against a wall, 3) Sitting behind a table and 4) Sitting in the middle of a room. Each subject also experienced a series of videos showing the same two (from the four possible) scenarios, with an actor taking the place of the subject. The videos were edited in such a way as to show a third person view of the room to set the scene and show an overview of the action, cutting to a close first person view as the robot came close to the subject. The order of the video and live trials was randomised for each trial, as was the order that each subject experienced the sequence of approach directions for both video and live trials. It has not been yet been possible to do exhaustive analysis of the results from these trials, but preliminary indications are that:

- The findings from the previous pilot study with 15 subjects have been largely confirmed and extended to a larger set of robot approach contexts.
- Most subjects disliked a direct frontal approach, even when behind a table; though a small proportion of female subjects actually preferred a frontal approach.
- Most subjects did not like being approached from behind by the robot.
- There was no clear preference overall for a left or right close approach.

An in-depth analysis of the results will be completed in 2006. With regard to the comparison of Video vs. Live trial methods, provisional analysis indicates that for certain types of scenarios and trials, video based HRI methods can be used to provide a large sample set in far less time than for live trials. The video based trials, while not having the same robot presence as live trials, can provide a quicker, cheaper and easier method of gaining a large number of results, which can be used to guide future full-scale live confirmatory trials. Further analysis of the results and data obtained will be carried out in the early months of 2006.

Figure 1. The Robot to Human Approach Trials: A robot approaches a sitting subject.

Work Package 3.2: Posture and Positioning

This workpackage, lead by Kungliga Tekniska Högskolan (KTH), on the role of posture and positioning in task-oriented Human-Robot Interaction is based upon user studies that provide data on how robot users and the robot coordinate their actions in a shared space. Related to the CF-SOC, “Socially acceptable interaction with regard to spaces” and the CF-DLG, “Multimodal dialogue including robust speech understanding for natural communication”, the work is targeting Key Experiment 1 (KE1) termed “The Home Tour”.

During the first period an initial HRI user trial was carried out with 22 subjects who showed a robot around in a living-room like environment, teaching it objects and places. Once an object or a place was learnt, the user could send the robot there again as a validation. The experiment was conducted with the Wizard-of-Oz methodology, and the robot was tele-operated both in navigation and multimodal interaction. The analysis of the collected data was informed and structured by the theoretical framework of Hall's Proxemics on interaction distances and Kendon's F-formation arrangements [Hüttenrauch et al. 2005].

WP3.2: Research Activities

During phase II on the study on posture and positioning two main activities were performed: The previously carried out HRI trial was quantitatively and qualitatively analyzed in selected parts. Additionally, a new HRI trial (termed "Multiple Room HRI-trial") according to the same KE1 scenario was initiated as a pilot study in cooperation with WP5.1. This new user trial incorporates the joint human-robot traversing of multiple rooms, with an integrated system tracking and following the user.

Findings from the initial HRI trial can be differentiated according to their quantitative and qualitative nature. Furthermore, as this analysis is based upon processing, transcribing, and annotating the HRI trial data, a data-rich HRI corpus of annotated video, still-images, and transcriptions was generated as well.

WP3.2: HRI Trial Findings

The HRI trial was set-up according to the KE1 of the "Home Tour" in which a user navigates a robot around and teaches it relevant places and objects to be later handled in robot-service missions. According to the different tasks performed these different HRI tasks were termed "Follow", "Show", and "Validate".

A quantitative analysis was performed on the interaction initiations of these interaction sequences of "Follow-", "Show-", and "Validate-" to investigate the spatial distance and F-formation arrangements between a human and a robot [Hüttenrauch et al., 2005].

It was found that Hall's "personal zone" is dominant when initiating the interaction with the used PeopleBot, i.e. independent upon mission –type subjects preferred to position themselves in the range of 1 to 4 feet (or ~.45 to ~1.2 meters). Furthermore, even the Kendon F-formation arrangements signifying the orientation and potential for joint action showed a dominant pattern: The "Vis-à-vis" (or face-to-face) positioning of the user towards the robot was found to be the most common formation not only in "Follow", but also in "Show", and "Validate" mission initiations. In other words the "L-Shape" F-formation, facilitating especially the interaction between two participants with a shared object in-between them, was in absolute terms less often encountered than the Vis-à-vis positioning. However, this dominance of the Vis-à-vis formation is much weaker in the Show and Validate mission initiations where 39% of the Show and 31% of the Validate missions were started in a L-Shape formation. As a consequence both the Vis-à-vis as well as the L-Shape formation should be anticipated and designed for in the spatial management behaviours of a robot companion.

Another finding regards the *dynamic* aspects in the spatial management for interaction: It could be shown from Laser range finder plots that regular patterns of interaction distance characterize the "Follow" as well as the "Show" episodes of interaction. Furthermore, the actual beginning and end of an interaction sequence are often marked by initial and ending *small position adaptations* on behalf of trial subjects. These observations, while not anticipated originally, are actually inline with Schegloff's descriptions [Schegloff, 1998]. In Schegloff's view embodied interaction sequences are guided by the potential of an initial "pre-departure" movement, a body posture during the ongoing interaction, and finally, a return or bodily changes into another posture and orientation during a single interaction sequence. These patterns of posture changes could be shown in the interaction with a robot, too.

The qualitative analysis of the interaction identified interesting research challenges that are relevant to be addressed with future robot behaviours; they are here only summarized briefly as bullet points:

1) **User approach and interaction initiation**. The robot's recognition and feedback signalling system to the approaching user is critical to avoid initial interaction and communication breakdowns.
2) **Joint-movement learning curve**. Spatial breakdowns (in the robot's failing to follow a user) are at least initially due to (a) the users' lack of experience of the robot's spatial behaviour, (b) the robot's missing anticipation of the user's inexperience, and finally, (c) the robot's following-behaviour that might fail due to numerous reasons To let a robot successfully follow a user, users need an initial training phase together with the robot to avoid (for example) too quick motions away from the robot, moving out of the robot's sensor systems "field of view" etc.
3) **Three-stages of spatial management**. Individual interaction sequences in multimodal dialogue as well as spatial management of posture and positioning might be modelled as a *preparatory*, a *current*, and a *transitional stage*. Identified as possible candidates for such a robot-based situation handling are for example proactive posture and orientation transition from "Follow" to "Show" episodes as well as the return form "Show" interaction sequences to a continued "robot-following" mode.
4) **Support for active user co-operation**. Multiple occurrences of users directly assisting the robot point towards the potential for users' willingness to act co-operatively and to "make the interaction work". Examples from the spatial posture and positioning observations include *placement of objects*, *getting-on-eye-level with robot, moving furniture, optimizing settings for the robot's mission*. Robots behaviours could make use of this HRI feature.
5) **Communicate robot's limitations**. Lack of prior HRI experience make the exact and initial differentiation between "can-be-done" vs. "does-not-work" difficult for users. Observations from our trials that show users' actively testing the robot's capabilities indicate a need (and user interest) in experiencing the true limit of a robot regarding navigation, speech dialogue interaction, object and gesture recognition, to name but a few.
6) **Allow users to shift between direct- and task-level robot control**. Task-level command and control of the robot was observed to lead to dead-locks both in spatial management between the robot and its users as well as in speech based dialogue. To allow users to fall back upon a more direct, tele-operation-like modus of commanding the robot, e.g. with a simplified keyword based command-language of robot navigation manoeuvres should be considered as a basis for future systems.

WP3.2: Multiple-Room HRI Trial

As an extension to the previously conducted HRI user study that reduced trial complexity by limiting the KE1 scenario of the "Home tour" to a single room, the new HRI study was expanded to include multiple rooms. Named "Multiple-room HRI trial" an initial explorative study was conducted in close co-operation with the KTH activity in RA5 on the "Spatial Cognition and multi-modal situation awareness" to be presented at HRI'06 [Topp et al., 2006].

In comparison to the first conducted study, implemented human tracking and robot navigation capabilities were used instead of simulated tele-operated movements of the robot. As another important extension the robot operation area was extended to incorporate a whole department floor (see figure 2) making it possible to investigate for example the joint human-robot passing of narrow spaces.

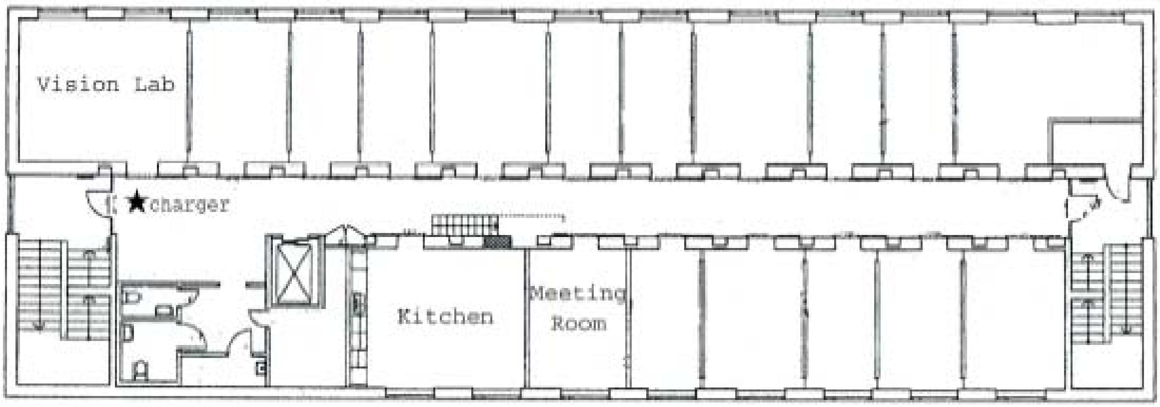


Figure 2. Office floor for the multiple-room HRI trial – The "star" marks the Robot's starting position during the trial

Contrary to the previous trial, the multiple-room HRI trial recruited subjects that are well familiar with the actual environment in which the trial was conducted. This change on the subjects' experience level was made due to the realization that it would be more valid for the KE1 interaction scenario of the "Home-Tour" that users are themselves well acquainted with the actual robot operation area. Unfamiliarity would instead lead to a need of teaching subjects the experiment-environment first, thus possibly biasing them with experiment instructions or the ad-hoc training of the spatial setting.

Successfully testing the multiple-room HRI interaction with 5 subjects, the detailed observations on posture and positioning are currently analyzed in order to prepare for a second extended part of this study that will be conducted and situated outside a research department environment if possible. A finding from this pilot study concerns the methodology and needs further investigation: Data collection in a multiple-room scenario is more complex and requires novel methods to observe and collect the spatial posture and positioning management between robots and their users.

Preliminary results [Topp et al. 2006] also showed that individuals apply many different strategies in teaching a robot places and locations in its future service environment. Individual preferences based upon personal appreciation of certain rooms and objects were found to play an important role in explaining these differences. Additionally, first indications for needed refinements in the human-robot joint traversing of the narrow spaces were observed, but require further analysis and empirical validation.

Work Package 3.3: Models and Algorithms for Motion in the Presence of Humans

The introduction of robots in our daily life raises a key issue that is "added" to the "standard challenge" of autonomous robots: the presence of humans in its environment and the necessity to interact with them. Clearly, the human should be taken explicitly into account in all steps of the robot design. In WP3.3 we aim to develop a motion and manipulation planner that explicitly takes into account the human by reasoning about his accessibility, his vision field and potential shared motions.

In the first year of the project, we had done an extensive literature research on properties of human-robot interactions and existing approaches to model these interactions. We had seen that a robot co-habiting with humans must have a number of physical properties like round joints, soft materials and friendly shape, and also intellectual properties like behaving more friendly by taking into account humans' comfort and reasoning about humans' properties.

In this deliverable we show the first results of the "Human Aware Motion Planner" that we have developed. The technical details of the planner are summarized briefly. More details about the inner workings of the system can be found in the attached papers [Sisbot2005].

WP3.3: Human Aware Motion Planner (HAMP)

User studies with humans [Dautenhahn et al., 2006] and robots provided us a number of constraints for this type of interaction. We present a new method that allows the integration of such additional constraints in a generic way. We defined three criteria that must be taken into account at the planning stage respecting the common aspects of motion planning such as obstacle avoidance. Each Criterion is presented by a 2D grid containing various numerical costs:

- **Safety Criterion**: The robot must avoid to collide with the human in the environment, and if possible it must prefer to not to pass too close to him. This criterion is represented by a human centred bell shaped grid. The radius of the bell and the costs values depend on humans state, as in figure 3 we can see, for example, that the safety zone tends to be smaller when the human is standing than sitting.

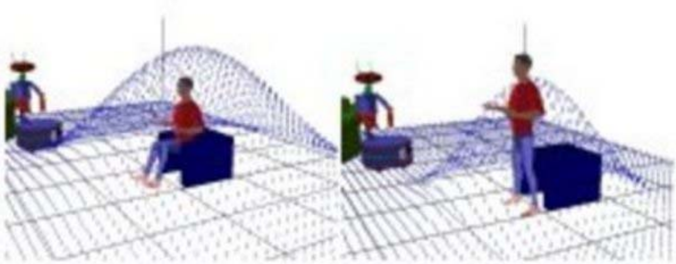


Figure 3. Costs in Safety Grid depends on humans state, like sitting or standing

- **Visibility Criterion**: The robot, if possible, should prefer to stay in the human's field of view. This criterion takes into account the humans visual field and allows us to produce more comfortable robot motions. The visibility grid (figure 4) is constructed according to costs reflecting the effort required by the human to get the robot in his field of view. Grid points located in a direction for which the human has only to move his eyes have a lower cost than positions requiring to move the head in order to get the robot in the field of view. Also, when the robot is far away from the human, the effect of the visibility must decrease.

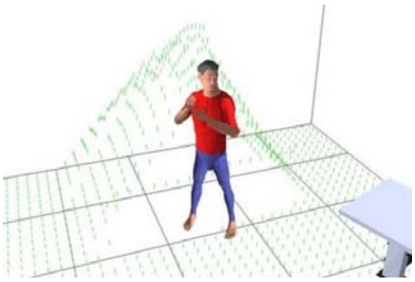


Figure 4. Costs in Visibility Grid depends on the humans field of view

- **Hidden Zones**: The robot, if possible, must avoid bursting out close to the human to avoid causing fear and surprise. This can be done by putting costs to the zones hidden by the obstacles. The costs in the hidden zone grid (figure 5) are inversely proportional to the distance between the human and the robot.

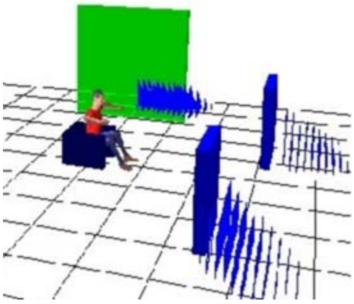


Figure 5. Hidden Zones are discouraged for the robot to cross over

These grids are combined to form a final grid in which we find a minimum-cost trajectory for the robot by using A* search algorithm. The found path is not only collision free and feasible but also human friendly with satisfying all three criteria. This method allows us to reflect each human's preferences as input parameters of grids giving us a level of flexibility on dealing different human behaviours. Indeed, it must be noted that while these different grids correspond to distinct criteria involving specific computation, the actual values can be tuned according to situations or to specific cultural habits.

WP3.3: An Example with Results

To illustrate the features of our planner, we can consider the scenario illustrated in figure 6 representing a living room and a kitchen in presence of two persons, Clark and Bruce, in the environment. We will look at the synthesised trajectories between the living room and the kitchen for different possible situations.

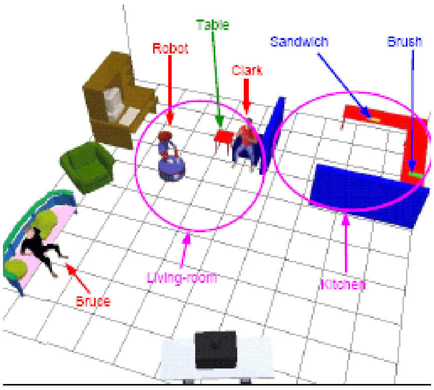


Figure 6. A living room scenario with 2 humans and a robot in the environment

In figure 7a, we imagine a situation in which Clark orders the robot to bring something from the kitchen. We see the path generated by the motion planner. The plan produced by HAMP takes into account the safety and the comfort of both humans by trying to stay in the visibility fields. We can see in figure 7b that HAMP finds a path that avoids bursting from behind the kitchen wall and so causing discomfort. The robot chooses a path that keeps a certain distance to this wall. In figure 7c, we can see that Bruce came to talk to Clark, so the robot calculates a different trajectory which maintains the visibility of Clark and also avoids passing too near to Bruce's back. As we use minimum cost approach in HAMP, if the path is blocked by an obstacle or a person, the robot chooses an alternative path (Figure 7d).

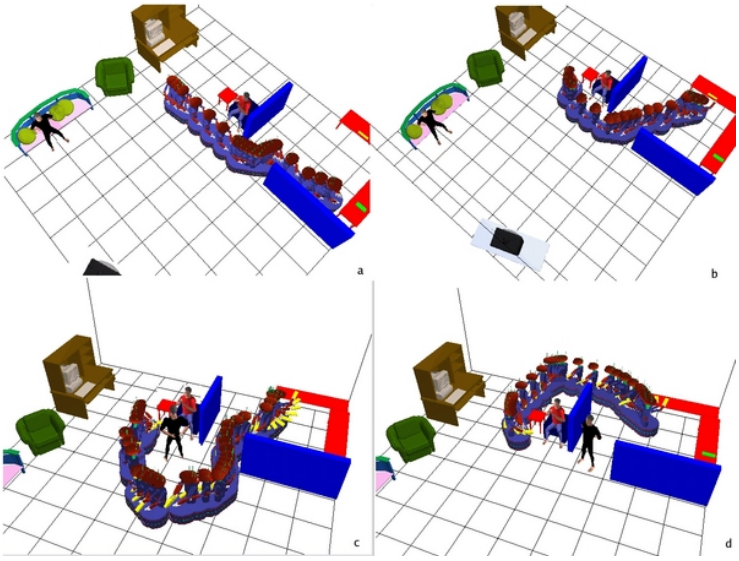


Figure 7. Four different situations results different paths in the same environment

Our planner is fast enough to re-plan and adapt its path along the execution. If a grid modifying change occurs, like a change in human state or position, or appearance of a dynamic obstacle, the computation times allows us to re-plan on-line and pass to the new path smoothly. Table 1 shows the processing times taken by the planning system for the examples above.

Grid Resolution	**Figure 7a**	**Figure 7b**	**Figure 7c**	**Figure 7d**
0.2 m	0.07s	0.09s	0.06s	0.15s
0.1 m	0.21s	0.25s	0.23s	0.50s
0.05 m	0.44s	0.78s	0.49s	0.20s

Table 1. Processor times that the planner took for the four examples above

WP3.3: Conclusion and Future Works

We have described the working mechanism along with an example and some results. Simulation results are encouraging and as our planner will be used in KE2, we are currently implementing it on our robot (Rackham) to have results from our planner for real situations. Another tool that we have is the proactive path planning [Krishna 2006] that allows the robot create safe paths by taking into account the possible hidden dynamic obstacles. This approach guarantees safe and “comfortable” paths when no human is in the environment. A combination of the proactive path planner and human aware planner will provide more reactivity by sharing the path to the two planners and also safe and comfortable motion by avoiding unseen obstacles or humans.

Manipulation in the presence of humans is also an important aspect when humans and robots need to interact closely to exchange objects or manipulating a common object. One part of our future work is to plan motions for manipulation tasks. The existence of humans in the task area also brings new concepts to the manipulation planning. Not only must the robot's motion be calculated, taking into account the human’s dynamics and kinematics, but also the robot's placement must take into account the human’s comfort. In Figure 8 we can see two uncomfortable robot placements for an interaction and in Figure 9 we see a suitable robot placement that allows a friendly interaction.

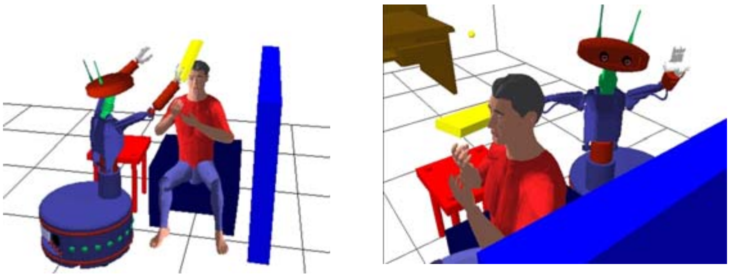


Figure 8. Two “bad” placements

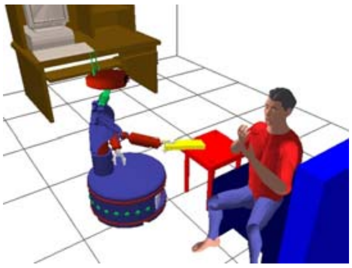


Figure 9. A suitable placement

Work Package 3.4: Requirements for Contextual Interpretation of Body Postures and Human Activities

Note, since the work carried out in WP3.4 has not been published yet, it is described in more detail than the other WPs. The lead partner for this WP is UH.

WP3.4: Naturally occurring gestures in a human-robot teaching scenario

Within WP3.4 (starting in June 2005) UH has been investigating gestures to facilitate the communication and interaction between humans and robots. During September/October 2005 user studies were performed to illuminate which naturally occurring gestures can be observed in a scenario specifically relevant to KE3. The term ‘naturally’ here refers to an unconstrained scenario where subjects were not given any scripts or pre-defined gestures to use. An important task up to now has been the development of a coding scheme to classify gestures people produce when asked to demonstrate how to perform a task. The coding scheme is an essential part of our strategy to systematically study the frequency, duration and sequence of different gestures in people's task demonstrations. In turn, this analysis will inform the design and parameterisation of algorithms for the recognition of human activities. This work was inspired by previous research in COGNIRON, specifically: Nehaniv et al. [2005] provided the conceptual framework for the coding scheme categories and research by University of Karlsruhe [2004] informed our present work in relation to requisites for human activity descriptions from a system's perspective.

The line of work described here is being pursued in close partnership with University of Karlsruhe (partner UniKarl) and their research within RA2 and KE3. A two-day workshop held at UH in October 2005 sought to elicit requirements for a coding scheme useful for the automatic classification of human activities, especially gesture, with regard to their functional roles as relevant for human-robot interaction, e.g. in allowing the robot to recognize limited aspects of human intent in order to be able to respond appropriate. Such requirements include determining functional types and important features (e.g. duration distributions, components, role in interaction, etc.) of gestures in natural human activity (as opposed to ad hoc validation via informal testing by robot developers). Requirements for high-quality annotated corpora of gestures were related in discussion to technical requirements of existing and future robotic systems used in COGNIRON and beyond. The annotated corpus will be made available for inspection by all partners as a tool for informing research in this topic. Note, complementary to related work at KTH within WP3.4 (see below) in the context of KE1 ("Home Tour"), the work at UH focuses on KE3 (Skill learning), investigating situations where a human instructs/teaches a robot a task. In 2006 KTH and UH, as part of WP3.4, will compare the results achieved and discuss a consolidation of the work (cf. third implementation plan). A detailed account of the work carried out so far has been published as a University of Hertfordshire Internal Technical Report [Otero, 2006].

Briefly, research on multimodal interfaces and human-robot interaction seems to point out the following approaches for the inclusion of gestures in the communication process; Some research explored the use of sets of pre-defined gestures and speech to communicate with robots [Ghidary, Nakata, Saito, Hattori, & Takamori, 2002; Oh, Lee, & You, 2005; Severinson-Eklundh, Green, & Hüttenrauch, 2003]. Other researchers consider that the use of gestures for communication with computer artefacts can and should be explored beyond the confines of a set of pre-defined gestures [Cassell, 1998, 2000; Cassell & Thorisson, 1999; Kipp, 2004].

WP3.4: Description of an initial exploratory study

This initial and exploratory study aimed at finding out to what extent robots' appearance influence the way people use different gestures to explain a certain task to a robot. The robot's appearance was divided into three categories: (a) human like robots, (b) animal like robots and (c) mechanistic look robots. The experiment had two phases. In the first one, participants were asked to sort a series of robots' photographs into distinct groups in order to assess to what extent the initial categories of the robots' appearance were relevant. In the second phase, participants were asked to instruct a robot on how to perform a certain task using only gestures and also to actually demonstrate how it should be completed. The participants were asked to do the experimental task for the three types of robots considered above in front of a video camera and imagine that this was the vision system of the robot. A software program simulated the robot's understanding of the gestures produced by the participants. The feedback consisted of the display of three colours in a computer screen. One of the experimenters was controlling the display of the feedback by pressing a button at the end of each sequence of the participants' gestures. The participants were prompted to consider the three different types of robots using the sorted photographs previously utilised. The session finished with semi-structured individual interviews.

WP3.4: Summary of results

Ten of our participants came up with 25 different designations for the groups of photographs they formed. From this initial number of designations we were able to aggregate into four groups: human-like, mechanistic-like, animal-like and toys. The participants agreed fairly well on which robots' photographs to include in the groups they formed. However, two photographs were more difficult to classify. The photograph showing robot Care-O-Bot (by IPA) was classified either as humanoid or mechanistic and the photograph showing the robot EMIEW (by Hitachi, Ltd) was classified as toy or humanoid. After analysis of the video recordings, the results suggest:

- Two participants who used more referencing with respect to objects also used more referencing with respect to places. The other participants used more miming manipulation and miming transportation. This indicates two different strategies: use of referencing or use of miming for the demonstrations.

- The gestures coded mostly last for 1 to 2 seconds. Frequencies for gestures lasting more than 3 seconds are lower. Some occurrences were of longer duration, especially when participants were asked to repeat their demonstration.

- When participants repeated their demonstration, they followed the same strategy through the session since the changes made mostly involve duration.

- The coding of more than one type of behaviour at the same time interval (one second) did occur - we termed these multiple occurrences. However, our present coding scheme was not sufficiently detailed to provide a clear picture of the different types of multiple occurrences. Nevertheless, the analysis suggests that the co-occurrence of referencing and miming as well as the occurrence of non-congruent eye gaze with the target of miming behaviour (for example, looking at a place while, at the same time, picking an object placed in a different location) are candidates for a more detailed analysis.

Note, the task set up for the second phase of the experiment was simple and it was expected that the recognition of the gestures' meaning should not be difficult but the observers/coders had some trouble coding the gestures. We need to improve the description of the different categories in order to improve inter-observer/coder agreement. In support of the tentative nature of this first study, some research points out the difficulties subjects have of unambiguously interpreting gestures alone [Hadar & Pinchas-Zamir, 2004]. In interviews conducted with the subjects, we found:

- The preferences were split between those for humanoid-like and animal-like robots for home use. None preferred the mechanistic-like robot.

- The participants seem to see themselves doing a similar task (teaching a robot using gestures) in their real life.

- Ten of the participants stated that they would be willing to go through a manual concerning a set of pre-defined gestures in order to communicate with the robot. When asked which system they would prefer, 6 considered the teaching and use of self-defined gestures, while 4 mentioned utilising a manual with pre-defined gestures.

- When asked if they were influenced by the robots' appearance, 6 participants were affirmative while 3 stated that the outer appearance should not make a difference. However, from the analysis of the videos, the categories used in the coding scheme did not seem to reveal any differences in the gestures produced. Nevertheless, the coders could perceive some subtle differences, in some cases, regarding the way the movements were performed: with mechanistic robots people tended to produce more jerky movements in comparison to humanoid robots where people's movements seemed more smooth.

- Finally, 5 of the participants were particularly aware of the difference between the animal-like robots and the other two types... Some statements: "I couldn't figure out how the biologically inspired robots could reproduce my gestures."; "...for the humanoid I should be able to just point."

Our initial study focused on the production of gestures alone. On the one hand, it seems reasonable to consider that people may not find "natural" gestures alone for the explanation of tasks to a robot. On the other hand, research points out that people's explanations using gestures alone of assembling tasks seem to facilitate not only their comprehension of the task but also the comprehension of the observers (recipients) of their explanations [Lozano & Tversky, 2004, 2005]. Hence, further research is needed to understand the roles gesture and speech might assume in human-robot interaction. Future work will contribute to integrating the advances in the understanding of gestures in HRI onto on-board recognition systems, and in the longer term, eventually in behaviour generation systems for interactive social robots - cf. Nehaniv et al. [2005].

WP 3.4: Development of an annotated video corpus in a robot-human following scenario

During the second half of 2005 discussions have taken place between the COGNIRON teams of University of Hertfordshire (UH) and University of Amsterdam (UvA) regarding the development of an annotated video corpus with the primary purpose of obtaining annotated video footage of a robot following human subjects. An experimental schedule has been developed for initial user trials to be carried out jointly by UH and UvA early in 2006. The prime video source is to be obtained from a robot mounted 360 degree camera as used by UvA on their robot platform. It is intended to annotate the video with an indication as to the experienced comfort of the following distance (according to the subjects' judgement) in terms of too close, too far, and just right (using a Comfort Level Device, CLD, developed by UH). An initial study will test the experimental set up and methodology and identify important issues for a further in-depth study planned later in 2006. This work contributes to WP3.4, with particular relevance also to RA 1 and 2 and Key Experiments (KEs) 1 and 3.

WP 3.4: Development of a taxonomy of body postures and human activities

Departing from the experimental data collected in the first phase, KTH has begun work on a taxonomy of postures and body movements relevant to KE1, using the human-robot dialogue as a guiding context. The taxonomy will have a strong empirical focus but also relate to general principles of human communication. KTH has worked with two parallel foci in the initial phase of WP3.4:

- Reviewing literature of existing taxonomies and related theories of communication
- Corpus development in relation to the work done in RA1 [Green et al, submitted], primarily based on the Home Tour scenario related to Key Experiment 1 (see D1.4.1).

The consequence of the empirical focus is that we will proceed in a bottom-up manner, by classifying instances of body movements (e.g., gesture, posture), spatial arrangements of agents (e.g. positioning), partly together with verbal communicative acts.

The other perspective we take in this work, which can be characterized as being top-down, is to relate the taxonomy to relevant theories on natural language communication. Initially we have reviewed taxonomies and approaches for classification of body movement and communicative behaviour. Furthermore we assume that classification of human-robot communication should be a *bilateral account* [Clark 2004], following work on human-human communication. Accordingly body movements and gestures are interpreted on basis of what the interaction participants aim to achieve by issuing a specific gesture. Hence approaches that consider human-human communication a joint activity (e.g. [Clark 1996, Clark 2004, Schegloff 1998, Gill 2000, Kendon 1990]) are relevant to our effort together with research on embodiment (e.g. Goodwin [2000]) and communicative context [Bunt 2000]. In bilateral accounts of communicative processes the focus of attention of the participants determines what is interpreted as context at a particular moment during interaction. The concept of relevance [Grice 1975] is also strongly related to context and the implications of relevant input in the light of the current context [Wilson 2004].

Developing a corpus and relating it to taxonomy are two activities that are conceptually interleaved. Given the work performed so far in RA1 (WP1.3) and RA3 (WP3.2) and findings research that has been surveyed, our initial conclusions are that taxonomy of human-robot communication should at least take these dimensions into account:

- A functional account for the intended (physical and mental) effects of a bodily movement in participants, and the immediate context of the participants, e.g., reference, attention, interaction space, etc.
- An account of morphological aspects of the configuration of participants and context, their movement dynamics, e.g., gesture displays, positioning movements, posture changes, etc.

RA3 Future work

Please consult the new implementation plan for phase3 and phase4 which details the future work to be carried out within RA3 in 2006 and 2007. Generally, the main themes for future work are:

- Preserving the structure of RA3 including the continuation of all four WPs
- Scientifically, continuing the main research issues addressed in the individual WPs with a focus on in-depth studies
- Increased integration with other RAs and partners in the light of KEs and increased joint research with partners from within RA3 and from other RA’s (leading to joint publications)

Specifically, WP3.1 will continue research on “social rules” for implementing socially acceptable, non-verbal behaviour in robots with respect to social distances in human-robot and robot-human approach scenarios, including handing over an object. The experimental scenarios will focus on naturalistic settings and long-term human-robot interactions. WP3.2 will address the role of posture and positioning in task-oriented robot-human interaction, based on user studies that will provide data on how robot and human coordinate their actions in a shared space, including a multiple room scenario, where spatial coordination is necessary for natural communication between human and robot. Work within WP3.3 will continue to develop models and algorithms for motion in presence and in the vicinity of humans. The approach is to endow the robot with a planner that will allow it to assess the feasibility of the motion and manipulation tasks it has to achieve in presence and in the vicinity of humans and/or in close coordination with humans, to share the load between the robot and the human and to explain/illustrate (when necessary) a possible course of actions. WP3.4 experimentally investigates which gestures and body postures/movements 'naturally' occur (spontaneously, without providing a predefined set) in situations such as described in the KEs. Research within WP3.4 will continue and expand during project phase 3 and 4, including a) further in depth analysis of work done in 2005, b) investigate consolidation of taxonomies developed at UH (studies relevant to KE3) and KTH (studies relevant to KE1).

References

COGNIRON RA3 Partners Publications

Note: These papers have been produced by the COGNIRON partners in the course of their work for RA3. For the reviewers' convenience they are included in the Appendix where final versions are available. In most cases they can be also be downloaded directly from the respective COGNIRON partners' websites. Details are available from the COGNIRON project internal website at www.cogniron.org.publications.php.

R. te Boekhorst, M. L. Walters, K. L. Koay, K. Dautenhahn, C. Nehaniv (2005). A Study of a Single Robot Interacting with Groups of Children in a Rotation Game Scenario, *Proc. IEEE International Symposium on Computational Intelligence in Robotics and Automation* (CIRA 2005), June 27-30, 2005, Espoo, Finland, pp. 35-40.

K. Dautenhahn, M. Walters, S. Woods, K. L. Koay, C. Nehaniv, E. Sisbot, R. Alami, T. Simeon (2006). How may I serve you? A Robot Companion Approaching a Seated Person in a Helping Context. To appear in *Proc. 1st Annual Conference on Human-Robot Interaction* (HRI'06), Salt Lake City, Utah, USA,

Green, A, Hüttenrauch, H, Topp, E and Severinson Eklundh K (2006). Developing a Contextualized Multimodal Corpus for Human-Robot Interaction, Submitted to LREC 2006: *5th Conference on Language Resources and Evaluation*, Genoa, Italy.

K. Madhava Krishna, R. Alami, T. Simeon , Safe proactive plans and their execution, *Robotics and Automation System,* In press, 2006

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

C. L. Nehaniv (2005). Classifying Gestures and Inferring Intent. *Proc. AISB'05 Symposium on Robot Companions*: *Hard Problems and Open Challenges in Robot-Human Interaction*, The Society for the Study of Artificial Intelligence and Simulation of Behaviour, pp. 74-81.

E.A.Topp, H. Hüttenrauch, H.I. Christensen, K. Severinson Eklundh (2006). Acquiring a Shared Environment Representation. To appear in *Proc. 1st Annual Conference on Human-Robot Interaction* (HRI'06), Salt Lake City, Utah, March 2006.

E. A. Sisbot, R. Alami, T. Simeon, K. Dautenhahn, M. L. Walters, S. N. Woods, K. L. Koay, C. L. Nehaniv (2005). Navigation in the presence of humans. *Proc. IEEE-RAS International Conference on Humanoid Robots* (Humanoids 2005), December 5-7, 2005 Tsukuba International Congress Center, Tsukuba, Japan. pp. 181-188.

M. L. Walters, K. Dautenhahn, K. L. Koay, C. Kaouri, R. te Boekhorst, C. L. Nehaniv, I. Werry, D. Lee (2005a). Close encounters: Spatial Distances Between People and a Robot of Mechanistic Appearance. *Proc. IEEE-RAS International Conference on Humanoid Robots* (Humanoids 2005), December 5-7, 2005 Tsukuba International Congress Center, Tsukuba, Japan, pp. 450-455

M. L. Walters, S. N. Woods, K. L. Koay, K. Dautenhahn (2005b). Practical and Methodological Challenges in Designing and Conducting Interaction Studies with Human Subjects. *Proc. AISB'05 Symposium on Robot Companions: Hard Problems and Open Challenges in Human-Robot Interaction*, 14-15 April 2005, University of Hertfordshire, UK, pp. 110-120.

S. N. Woods, K. Dautenhahn, C. Kaouri (2005). Is Someone Watching Me? Consideration of Social Facilitation Effects in Human-Robot Interaction Experiments. *Proceedings 2005 IEEE International Symposium on Computational Intelligence in Robotics and Automation* (CIRA 2005), June 27-30, 2005, Espoo, Finland, pp. 53-60.

S.N.Woods, M L Walters, K L Koay, K Dautenhahn (2006). Comparing Human Robot Interaction Scenarios Using Live and Video Based Methods: Towards a Novel Methodological Approach. To appear in *Proc. IEEE Int. Workshop on AMC06* (Advanced Motion Control), Istanbul, March 27-29 2006.

N. Otero, C. Nehaniv, K. Dautenhahn, J. Saunders, A. Alissandrakis (2006). Naturally Occurring Gestures in a Human-Robot Teaching Scenario: An exploratory study. University of Hertfordshire, Science and Technology Research Institute, Computer Science Technical Report (443).

Hüttenrauch, H., Severinson Eklundh, K., Green, A., Topp, E.A. (2005). Investigating Spatial Relationships in Human-Robot Interaction. NADA, Technical Report.

Relevant COGNIRON Project Documents

University of Hertfordshire (2005). COGNIRON D 3.1.1. Report on Evaluation of user studies on interaction styles and personal spaces.

Kungliga Tekniska Högskolan - (2005). COGNIRON D 3.2.1. Report on User study on the role of posture and positioning in HRI.

University of Karlsruhe (2005). COGNIRON D.2.3.1. Report on a classification of human activities in household environments.

COGNIRON (2005a). D7.1.1 Specification of the 3 key experiments

COGNIRON (2005b). D7.2.1 Set up of the Key Experiment “Robot Home Tour”

COGNIRON (2005c). D7.3.1 Set up of the Key Experiment “Curious Robot”

Other References

Cassell, J. (1998). A Framework For Gesture Generation And Interpretation. In R. Cipolla & A. Pentland (Eds.), *Computer Vision in Human-Machine Interaction* (pp. 191-215). New York: Cambridge University Press.

Cassell, J. (2000). Nudge Nudge Wink Wink: Elements of Face-to-Face Conversation for Embodied Conversational Agents. In J. Cassell & J. Sullivan & S. Prevost & E. Churchill

(Eds.), *Embodied Conversational Agents* (pp. 1-27). Cambridge, Massachusetts: The MIT Press.

Cassell, J., & Thorisson, K. R. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, 13(4-5), 519-538.

Ghidary, S., Nakata, Y., Saito, H., Hattori, M., & Takamori, T. (2002). Multi-Modal Interaction of Human and Home Robot in the Context of Room Map Generation. *Autonomous Robots*, 13(2), 169-184.

Hadar, U., & Pinchas-Zamir, L. (2004). The semantic specificity of gesture - Implications for gesture classification and function. *Journal of Language And Social Psychology*, 23(2), 204-214.

Kipp, M. (2004). Gesture Generation by Imitation: From Human Behaviour to Computer Character Animation. Boca Raton, Florida: Dissertation.com.

Lozano, S., & Tversky, B. (2004). Communicative Gestures Benefit Communicators. In K. Forbus & D. Gentner & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 849-854). New Jersey: Lawrence Erlbaum Associates.

Schegloff, E. A., 'Body torque', *Social Research*, 65, 535 – 586 (1998).

Lozano, S., & Tversky, B. (2005). Gestures Convey Semantic Content for Self and Others. In B. Bara & L. Barsalou & M. Bucciarelli (Eds.), *Proceedings of the Twenty Seventh Annual Conference of the Cognitive Science Society* (pp. 1337-1342). New Jersey: Lawrence Erlbaum Associates.

Oh, J. Y., Lee, C. W., & You, B. J. (2005). Gesture recognition by attention control method for intelligent humanoid robot, *Knowledge-Based Intelligent Information and Engineering Systems, Pt 1, Proceedings* (Vol. 3681, pp. 1139-1145).

Severinson-Eklundh, K., Green, A., & Huttenrauch, H. (2003). Social and collaborative aspects of interaction with a service robot. *Robotics and Autonomous Systems*, 42(3-4), 223.

Bunt, H.C (2000) Dialogue pragmatics and context specification. Harry Bunt and William Black (eds) Abduction, Belief and Context in Dialogue. *Studies in Computational Pragmatics*. John Benjamins, Amsterdam, 2000, Series Natural Language Processing, Volume 1, pp. 81-150.

Clark H.H. (1996) *Using language*, Cambridge University Press.

Clark, H.H and Krych. M (2004) Speaking while monitoring addressees for understanding, *Journal of Memory and Language* 50 (2004) 62–8.

Goodwin. C (2000). Action and Embodiment Within Situated Human Interaction. *Journal of Pragmatics* 32: 1489-522.

Grice, J. P. (1975) Logic and conversation In P. Cole and J. L. Morgan, editors, *Syntax and Semantics*, volume 3: Speech Acts, pages 41–58. Academic Press, New York, NY, 1975.

Kendon, A. (1990). *Conducting interaction - Patterns of behavior in focused encounters. Studies in interactional sociolinguistics*. Cambridge, NY, USA: Press syndicate of the University of Cambridge

Schegloff E.A (1998) Body Torque. *Social Research*, 65:3, 1998, 535-596.
Wilson, D & Sperber, D (2004) Relevance Theory. In (Horn, L.R. & Ward, G. eds.) *The Handbook of Pragmatics*. Oxford: Blackwell, 607-632.

Appendices

A Study of a Single Robot Interacting with Groups of Children in a Rotation Game Scenario[*]

René te Boekhorst, Michael L. Walters, Kheng Lee Koay, Kerstin Dautenhahn , Chrystopher L. Nehaniv
{R.teBoekhorst, M.L.Walters, K.L.Koay, K.Dautenhahn, C.L.Nehaniv} @herts.ac.uk

Adaptive Systems Research Group
School of Computer Science
University of Hertfordshire

College Lane, Hatfield, Herts AL10 9AB, United Kingdom

Abstract **- We tested the hypothesis that children are more attentive to a robot if the robot appears to be interested in the children. In addition, we investigated if and how the quality and quantity of a child's attentive behaviour varies with the distance to the robot, reflecting the notion of "social spaces". Hereto, 16 groups of up to 10 children each were engaged in a play scenario in which they had to move closer to a robot over 6 successive rounds. The robot was endowed with a "searching" camera "eye" and a pointing arm and hand that could either be fixed in a static position ("static pointing") or actively rise to point at a child ("active pointing"). The results showed that:**

1) The mean frequency of overall attentive behaviour (including towards other children) was significantly higher when the robot was not pointing at the children and independent of the state of the camera.

2) "Looking at" was the most frequently scored attentive activity for the children and was mainly targeted to the robot, but not correlated with any of their other attentive activities.

3) There was an interaction effect between the state of the camera and of the pointer: looking at the robot by the children occurred significantly more often when the camera and the arm were consistent in signalling apparent interest (i.e. both active or both passive).

4) There was no demonstrable effect of distance to the robot on the overall attentive behaviour of the children.

Index Terms - Human-Robot Interaction, Social Spaces, Social Interaction, Social Distances, Social Robot, Deixis.

I. INTRODUCTION

The impact of design features on the attraction and acceptance of agents by humans is a major topic in Human-Agent Interaction research [Breazeal 2002, Dautenhahn et al. 2002, Dryer 1999]. Acceptance hinges on feelings of control and comfort [Norman 1994]. Whereas control is about having information (knowing how an agent can be operated) to enable manipulative and corrective operations, comfort has to do with the social aspects of acceptability, such as trust. Trust in the agent, in turn, is related to the reassurance that all is working to plan and therefore requires the actions of the agent to be understandable (knowing what an agent does) [Norman, 1997]. Critical for the latter is that an agent makes its presence and actions known to the human. For this, attracting (and keeping) attention by the agent is of principal importance and the role of personification is frequently debated in this context (see for example Erickson [1997]; Takeuchi et al. [1995]). A number of researchers assume that a believable personification is mainly obtained by an anthropomorphic shape. However, as will be argued in this paper, a believable and consistent combination of actions might be just as important (cf. [Dautenhahn 1998, Dautenhahn and Nehaniv 2000]).

The above issues have originally been raised for the design of software agents [Bradshaw, 1997], but they also directly apply to the design of hardware agents (robots). Apart from the role of attention in building up trust, a special topic that needs to be considered in the case of robots (but that does not apply to virtual agents) is the feeling of physical safety. In humans and animals this is related to the concept of "social spaces"; zones around the individual in which the type of interactions depend on their distance to that individual. Hall [1966] provided a basis for research into social and personal spaces between humans, and later work in psychology has demonstrated that social spatial distances and zones substantially reflect and influence social relationships and attitudes of people. Embodied non-verbal interactions, such as approach, touch, and avoidance, are fundamental to regulating human-human social interactions [Hall, 1968], and these insights have provided guidelines for more recent research, studies and investigations into human reactions to robots [Goetz and Kiesler 2002, Paiva et al. 2004, Scopelliti 2004, Woods, et al, 2004].

The study of attraction towards a robot on the one hand and feelings of comfort on the other hand (reflected by the association between attentive behaviour and social spaces) is of obvious importance to the emergent field of assistive robotics in general and to the design of robots for use in the home in particular (for an overview of socially interactive robots, see Fong et al. [2003]). As such, the investigation of this research question is part of our contribution to the COGNIRON project (a European Framework VI project [COGNIRON 2003, 2005]), of which the major objective is to develop a robot that serves humans as a companion in their daily lives.

---

[*] The work described in this paper was conducted within the EU Integrated Project COGNIRON ("The Cognitive Robot Companion") and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020.

To formulate the research question explicitly in the wider area of Human-Robot Interaction (HRI), we stress the notion of interaction by considering attention in the context of embodied interaction. Eye-gaze direction and deictic gestures such as pointing with the hand play important early roles in the ontogeny of inter-subjectivity and shared attention as well as linguistic and cognitive abilities in the development of human children; and they continue to play a crucial role in non-verbal communication in adults, depending crucially on embodied interaction [see, e.g. Kita 2003]. These considerations have lead to the following working hypothesis: humans are more attentive towards the robot if the robot appears to be "interested" in the people present.[†] We implemented this in two ways: First, the camera on top of the robot was engineered so that it could move around its axis and thus give the impression of being a searching "eye". Second, the robot was endowed with an arm and a human shaped "hand" that could rise to point to a particular, chosen person within a group of subjects. Experiments were carried out in which groups of up to 10 school children were confronted with such a robot in a play scenario. The behaviour of the children was scored at various distances to the robot and compared under the four combined conditions of camera or pointer being static or active.

Fig. 1: The PeopleBot™ robots fitted with a pointer (in raised and lowered positions) and basket.

Fig. 2: Extent of the robot's camera movement

II. METHODS

The study took place in June 2004, taking advantage of a larger event run by the FP5 European Project VICTEC [2003]. For the VICTEC project, about 400 school children (aged between 9 and 11 year and chosen from 10 schools in the Hertfordshire area) visited the university and 194 of these participated in the experimental sessions described in this paper. These sessions were used to run interactive games with groups of children and a single robot. The main aims of this study were to investigate interaction styles in a group scenario involving children and how they depend on proximity to the robot ("social spaces"). In particular, we wanted to find out how a robot can attract the attention (by using combinations of two attention attraction devices) of humans and how humans react to being the focus of the robot's attention. The study was conducted using commercially available, human-scaled, PeopleBot robots [ActivMedia, 2005].

A. Experimental Set-up

The sessions took a maximum time of 30 minutes and were run in parallel in two separate rooms. Six sessions were carried out each day for four days. In each session a robot and a group of children were involved in an interactive "rotation game" consisting of six rounds. The two rooms to be used for the study were both 10m long by 6m wide. Each room contained a PeopleBot robot in the centre, initially covered by a black plastic sack. The room was marked out into 6 concentric zones around the robot at 0.5m radii. There was video recording equipment set up in each room consisting of two fixed video cameras, and a feed from the robot onboard video camera. The videos were time coded, for later synchronization and evaluation, and each day's sessions were recorded on a new set of tapes. The tapes were then downloaded and stored on computer readable compressed format (mpeg 1 or 2) on CD or DVD disk, for later evaluation. The robots were controlled in a semi-autonomous manner, with the operators retaining control of starting the games, and also making sure that the robots did not point to an empty space when selecting (by stopping the turning robot when it was facing) a child. In static pointing mode, the pointer did not move, but was constantly extended. In active pointing mode, the pointer was lowered most of the time, but raised when a child was selected. The two robot operators were hidden in an adjoining third room along with the wireless network, recording equipment, and the various data processing computers. The game (but not the purpose of the experiment) was explained to the children by an experimenter who initiated the game with a signal to the robot (seen by the robot operators). During the experiment, behaviour of the children was recorded by the video camera and another observer took notes of any particularly interesting interactions.

During the game, the robot revolved in the middle of a circle of children with various slow, fast, reverse and "teasing" movements to keep the children interested. After 30 seconds the robot stopped when it faced the nearest child and "selected" it by beeping twice, stopping in front of that child and – depending on the experimental condition – raising the pointer toward the child.

[†] We are not claiming that our robot was genuinely interested in the children, but refer to the usage of cues that give the appearance of "interest".

Fig. 3: A child being selected by the robot while playing the Rotation Game, with pointer rising to point to him (active pointer mode). Spatial zones are marked on floor.

The chosen child was allowed to pick up a small present from a basket carried by the robot and then left the game The remaining children moved 0.5m closer (to the next zone) to the robot and the game was repeated for a total of six rounds. At the end of the game, all the children that were left over received an identical present.

The passive pointer was always in the raised position pointing forward. The active pointer was normally in the lowered position (pointing to the floor) but when selecting a child the pointer was raised so that the hand was pointing with the finger at the person or object directly in front of the robot (Figs. 1 and 3) and then lowered once more before the child took the present.

In addition to an active pointing or a static pointing hand, the other variable condition was the state of the camera. The camera was either stationary or panning from side to side in a random manner (Fig. 2). Hence, the overall setup for statistical analysis is a 2 x 2 design of four experimental conditions (fig. 4):

1. Camera static and static pointing
2. Camera static and active pointing
3. Camera moving and static pointing
4. Camera moving and active pointing

The dependent variables are the frequencies of activities scored from a focal child (see section B. Collection, Processing and Analysis of the Data). Out of twenty-four rotation games, sixteen giving the best quality data were chosen for analysis. This resulted in a final data set consisting of four different groups for each of the four experimental conditions.

B. Collection, Processing and Analysis of the Data

From each videoed session, one child was selected for analysis, thus guaranteeing independence of data. Selection was based on the visibility of the child in the video clip and the duration of time he or she spent playing the game (i.e. children that dropped out of the game after the first two rounds were not considered). After preliminary viewing of the videos, (not necessarily robot-directed) "Pointing at", "Talking to" and "Looking at" were selected as behavioural activities for an initial analysis. These activities were understood as signs of interest and attention shown by the children to the environment (including the robot and other persons). Their summed frequency is considered as overall "attentive behaviour".

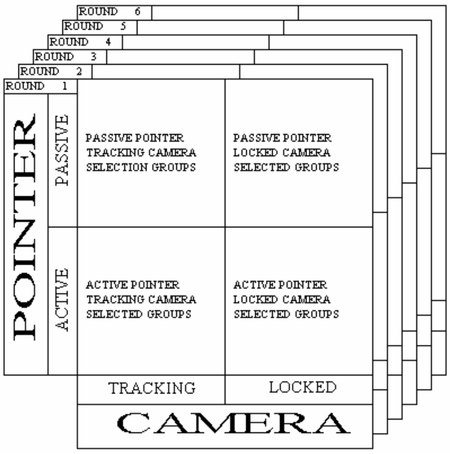


Fig. 4: The four experimental conditions, showing four combinations of active and static camera and pointer used for playing the rotation game with randomly selected groups of children. (Each lasted for six rounds, i.e. six children from each typical group of ten children were selected by the robot in the course of a game)

The selected activities were coded using the program ANNOTATOR developed by the University of Hertfordshire Adaptive Systems Research Group. ANNOTATOR enables the interactive recording of behavioural variables at fixed time intervals: coded activities can be scored and saved in a worksheet during the simultaneous display of a video clip of an experiment. This methodology of behavioural analysis has previously proved beneficial in analysing human-robot interaction [Dautenhahn and Werry 2002, Robins et al. 2004]. A one-second time unit was adopted for our purposes and attentive behavioural activities were scored according to the target they were directed at (Robot, Other Children, Other Persons - such as teacher or the instructing researcher - and "Unknown Target"). For the purpose of the analysis, attention targets coded as Other Person and Unknown Target were joined together into a lump-category "Else". The final data set thus consists of the frequency of the "Talking to", "Pointing at" and "Looking at" three targets (Robot, Other Children and Else) displayed by one child selected from each of four groups, giving a total of 16 data points.

The coding started from the beginning of the game to either the end of the game (when all remaining children were asked to take a present from the robot) or the end of the game for the child selected for coding (when he or she was picked by the robot). During the rotation game, the moment when children moved to the next circular zone was also coded (from the moment the child selected for coding had both feet in the next circle). Zones were coded from 1 (inner-most circle) to 5 (furthest from the robot).

We found a strong effect of the duration of the game on the frequency by which a subject displayed attentive behaviour. As a result of this, statistical tests could only be performed after controlling for the effect of duration or by making (matched) comparisons within subjects.

In the next section, we will respectively address: i) differences in overall attentive behaviour between the experimental conditions; ii) the allocation of directed attention; iii) the association between directed attention and the experimental conditions, and iv) the effect of proximity to the robot on overall attentive behaviour.

II. RESULTS

A. Differences in overall activity between experimental conditions.

The effect of time was taken into account by using the residuals from the regression of overall attentive activity (the sum of looking, talking and pointing) on the duration of an experiment. Residuals were computed for the activity scores of each subject and were found to meet the assumptions of analysis of variance (Bartlett's, Cox and Hartley statistics for homogeneity of variance; normality of error; no correlations between means and standard deviations of the samples).

A repeated two factor (State of Camera = moving, static, and State of the Pointer = active, static) ANOVA was carried out. A significant effect for the state of the hand was found ($F_{1,\ 12} = 5.73$, $p = 0.034$) and indicated that subjects exhibited more attentive behaviour (including towards each other) when the pointer was static as compared to the active pointer (Fig. 4).

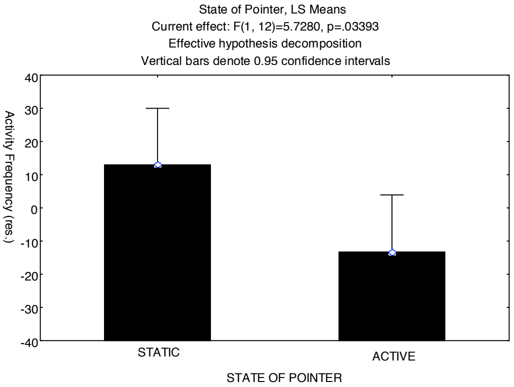


Fig. 4: The effect of the mode of the robot's pointing on the overall attentive behaviour of subjects (including attentiveness towards each other).

B. Allocation of Directed Attention

No significant effects were revealed for the state of the camera or for the interaction between the states of camera and pointer. As an alternative procedure an Analysis of Covariance was performed on the raw scores, but with Duration as co-variable. The results were identical to those reported above.

In the previous tests, only the overall attentive behaviour was considered. Table 1 shows how attentive behaviour was allocated by the subjects. "Looking at" was the most frequently scored activity and was mainly directed towards the robot. "Pointing at" and "Talking to" only made up for a small proportion of the overall activities and were in the majority of cases targeted towards the other children.

TABLE 1.
DISTRIBUTION OF THE DIRECTED ATTENTION VARIABLES

	DIRECTION			
ATTENTION VARIABLE	at robot	at children	else	Totals:
Looking	3066	1281	612	4959
Talking	59	164	5	228
Pointing	9	22	2	33
Totals:	3134	1467	619	5220

Some of the directed attention variables (after time-correction, i.e. the residuals from the regression on duration) appeared to be significantly correlated. For example pointing to the robot was significantly correlated with talking to the other children (Pearson correlation, $r = 0.59$). These correlations may be partly due to simultaneous attentive activity: when pointing to the robot, subjects may have talked about the robot at the same time with their group members. However, the reversed correlation (between talking to the robot and pointing at other children, $r = 0.57$) is less easy to explain as is the absence of correlations between talking to and pointing at the robot. The correlation between talking to the robot and pointing at their classmates could be due to attempts to influence the robot. E.g. a child would say "pick him!" to the robot, while pointing to their classmates. Furthermore, it is noteworthy that although "Looking at" was the dominant activity it appeared not to be correlated with any of the other directed attention variables. Further investigations need to be carried out to clarify these matters.

C. Associations between Directed Attention and Experimental Conditions.

In order to assess the children's attention towards the robot in a single measure, for each attention variable we calculated the "AttentionBy•X" variable (where X is one of the three attention variables "Looking at", "Talking to" or "Pointing at"):

$$\text{AttentionBy} \cdot \text{AttentionVariable} = \frac{\text{AttentionVariable} \cdot \text{robot} - \text{AttentionVariable} \cdot \text{notrobot}}{\text{AttentionVariable} \cdot \text{total}} \times 100$$

In this formula •notrobot refers to the frequency of that attention variable directed at children plus "else" (i.e. not targeted to the robot). Similarly, •total is the total frequency of the considered attention variable. For example, attention towards the robot by "Looking" at the robot is formulated as:

$$\text{AttentionBy} \cdot \text{Looking} = \frac{\text{Looking} \cdot \text{robot} - \text{Looking} \cdot \text{notrobot}}{\text{Looking} \cdot \text{total}} \times 100$$

The division by the total frequency makes this measure independent of duration. Each of the three "AttentionBy•X" variables was subjected to a two-factor (State of Camera, State of Pointer) ANOVA with replication.

The results show the significant effect of the interaction between State of Camera and State of Pointer (Fig. 5) for AttentionBy·Looking. More specifically, subjects paid more attention by looking at the robot when both pointer and camera were either moving or both were static. Children looked relatively less at the robot when camera and pointing were "conflicting" – one static and other not.

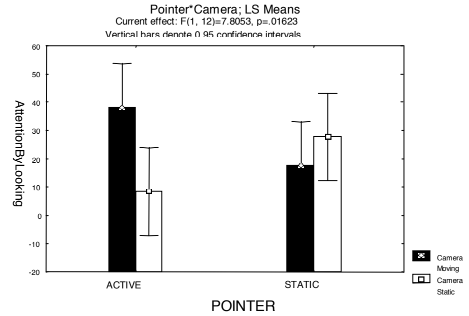


Fig. 5: Analysis of Variance for the variable AttentionBy•Looking under the two states of camera and pointer (active, static)

For AttentionBy•Pointing and AttentionBy•Talking no significant affects due to the camera or pointer state were found.

In order to examine in more detail the possible differences in attention children paid to the robot, other children and "else", within each experimental condition, a series of Friedman tests was performed. The Friedman test is a non-parametric two-way analysis of variance (without replication) for rank-ordered data. The test was applied to compare for each subject how often he/she displayed attention towards the three targets (robot, children, else = unknown + other persons). Friedman tests were carried out separately for each of the three attention variables ("Looking at", "Talking to", and "Pointing at") for each of the four experimental conditions (Table 2).

TAB LE 2.
RESULTS OF THE FRIEDMAN TESTS;
FOR EACH VARIABLE DIFFERENCES BETWEEN TARGETS WERE TESTED FOR SIGNIFICANCE

		LOOKING		**POINTING**		**TALKING**	
Camera	**Pointer**	χ^2	**p**	χ^2	**p**	χ^2	**p**
Moving	Active	6.5	0.04	NOT SIGNIFICANT		NOT SIGNIFICANT	
	Static	8	0.02				
Static	Active	12.25	0.002				
	Static	8	0.02			8	0.02

Pointing - No significant differences could be found in the frequency with which the children spent pointing towards the robot, other children or "else" within any of the experimental conditions. Pointing towards the other persons, a teacher or the researcher running the experiment did not occur at all.

Looking - For all experimental conditions we found significant differences in the frequency of "Looking at" the robot, other children or "else". In all cases children looked most often at the robot, less at other children and the least at other persons plus unknown targets.

Talking - Significant differences in how often children talked to the robot, other children or other persons/unknown targets were found only when the camera was static and the pointer was static: in that case children talked significantly more to other children than to the robot or "else" .

Although not significant, the same tendency was observed in the experimental condition 3 (camera moving, pointer static). Therefore, these two conditions were further analysed together. The outcome of a Friedman test performed for experimental conditions 1 and 3 together was significant (χ^2 = 12.074, df = 2, p = 0.002) and suggests that when the robot's pointer was static, irrespective of the camera status, children talked significantly more to other children than to the robot or to other persons/unknown targets.

D. The Effect of Proximity to the Robot

Similar analyses as described above were carried out taking the distance to the robot (i.e. the zones in which the subject was positioned in) into account. We did this by entering a code for zone (from 1 to 5, corresponding with closer to further distance) as an extra factor in the analysis of variance. However, this inclusion lead to serious violations of the assumptions underlying ANOVA and we therefore had to perform separate tests for each of the four experimental conditions. None of the tests showed significant effects.

II. CONCLUSIONS

The analysis showed that for all experimental conditions, irrespective the state of the camera or the pointer status, the highest scored attention variable was looking at the robot. The lack of significant differences in the frequency of pointing at the robot, other children or "else" might be explained by the rare occurrence of this behaviour during the experiment. Pointing at other persons other than the children did not occur at all, probably as a consequence of obeying social norms (especially in the U.K. this is regarded as rude and the "other person" in the experiment was always an adult). Alternatively, the "other persons" were not seen as involved in the game.

It is worth mentioning, that some children started mimicking the start 'thumbs up' signal, given to the robot by an experimenter at the beginning of each game. This could be interpreted as trying to influence robot's behaviour or to make a contact with it.

The seemingly "attentive" behaviour of the robot influenced the behaviour of the children in various ways, although not all of the effects are easy to interpret. Overall activity was highest when the pointer was static, which may be due to boredom, leading in turn to increased talking and interactions with the other children. The heightened attention by looking at the robot showed a surprising result: children were less interested when the camera and pointer were in conflicting states. Possibly this "confused" the children or was counter to their expectations of a "balanced" behaviour on the part of the robot. Apparently, human-robot interactions cannot be understood as one-dimensional responses to simple stimuli.

The only significant differences between talking to the robot, other children or other person/unknown targets were found for experimental condition 1 and for the combined data of condition 1 and 3 (static pointing). In both cases children more frequently talked to other children and less to the robot or “else”. Since in condition 1 the camera and pointer were both static, it is possible that the children interpreted this as a lack of interest from the side of the robot. This in turn might have made the children less interested in the robot and therefore led them to be engaged in a conversation among themselves instead. This is in line with our observation that the overall activity (with regard to the attentional behaviour), of the children was highest when the pointer was static. This fact and the outcomes of the Friedman test on the combined data from condition 1 and 3 (static pointing, independent of the state of the camera) suggest indeed that the robot’s active pointer was a stronger stimulus for the children than the camera. A possible explanation is that its movement was more conspicuous than that of the camera. Alternatively, the fact that the pointer’s shape included the form of a human hand might have directed the attention.

No effect of distance to the robot could yet be demonstrated, but at the time of writing only overall attentive behaviour was analysed. We are currently testing the influence of proximity for the separate directed attention variables.

ACKNOWLEDGMENTS

Thanks go to Christina Kaouri, Iain Werry, Ben Robins, and Aris Alissandrakis for help in carrying out the trails, to Dorota Jagielska for data recording and help in the statistical analysis, and to Alexander Klyubin was his work on the ANNOTATOR software and other technical assistance.

REFERENCES

ActivMedia (2005). Activmedia website with details of PeopleBots™ ; www.activmedia.com

Bradshaw, J. M. (1997). An introduction to software agents. In: *Software Agents*, ed. J. M. Bradshaw. Menlo Park, Calif. AAAI Press.

Breazeal, C.L., (2002). Designing Sociable Robots. Massachusetts: The MIT Press, 2002.

COGNIRON, (2003). Sixth Framework Programme Document Annex.

COGNIRON (2005). Report D6.2.1. Evaluation of User Studies on Intention and Attribution. Jan 2005.

Dautenhahn, K. (1998). The art of designing socially intelligent agents: Science, fiction, and the human in the loop. *Applied Artificial Intelligence*, 12(7-8): 573-617.

Dautenhahn, K. and C. L. Nehaniv, Living with Socially Intelligent Agents. In: K. Dautenhahn (ed.), *Human Cognition and Social Agent Technology*, John Benjamins Publ. Co., pp. 415-426.

Dautenhahn, K., B. Ogden, T. Quick (2002). From Embodied to Socially Embedded Agents - Implications for Interaction-Aware Robots, *Cognitive Systems Research* 3(3): 397-428.

Dautenhahn, K., I. Werry (2002) A Quantitative Technique for Analysing Robot-Human Interactions. *Proc. IROS2002, Lausanne, 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1132-1138, IEEE Press.

Dautenhahn, K. (2004). Robots We like to Live With?! – A – Developmental Perspective on a Personanlized Life-Long Robot Companion. Invited paper, *Proc. IEEE Ro-Man 2004, 13th Annual Workshop on Robot and Human Interactive communication, September 20-22, Kurashiki, Okayama, Japan.* IEEE Press, pp. 17-22.

Dryer, D. C. (1999). "Getting personal with computers: How to design personalities for agents.", *Applied Artificial Intelligence*, 13,:273-295, 1999.

Erickson, T. (1997). Designing agents as if people mattered. In: *Software Agents*, ed. J. M. Bradshaw. Menlo Park, Calif. AAAI Press.

Fong, T., I. Nourbakhsh, and K. Dautenhahn, (2003). A survey of socially interactive robots, *Robotics and Autonomous Systems*, 42: 143-166, 2003.

Goetz, J. and S. Kiesler, (2002). "Cooperation with a Robotic Assistant" presented at CHI'02 Conference on Human Factors in Computing Systems, New York, USA, 2002.

Hall, E. T. (1966). *The Hidden Dimension: Man's Use of Space in Public and Private*. The Bodley Head Ltd, London, UK. 1966

Hall, E. T. (1968). Proxemics. *Current Anthropology* 9(2-3): 83-108. 1968

Norman D. A., (1994). How might people interact with agents, *Communications of the ACM*, 37: 68-71.

Norman, D. A. (1997): How people might interact with agents? In: *Software Agents*, ed. J. M. Bradshaw. Menlo Park, Calif. AAAI Press.

Paiva A, J Dias, D Sobral. R Aylett, P Sobreperez, S Woods et al, (2004). Caring for Agents and Agents that Care: Building Empathic Relations with Synthetic Agents. *Third International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 1 (AAMAS'04),* pp. 194-201. July 19 - 23, 2004

Robins B., K. Dautenhahn, J. Dubowski, (2004). Investigating Autistic Children's Attitudes Towards Strangers with the Theatrical Robot - A New Experimental Paradigm in Human-Robot Interaction Studies, *Proc. IEEE Ro-man 2004*, IEEE Press, pp. 557-562.

Scopelliti M., M. V. Giuliani, A. M. D'Amico, and F. Fornara, (2004). "If I had a robot at home.... Peoples' representation of domestic robots," in *Designing a more inclusive world.* In: S. Keates, J. Clarkson, P. Langdon, and P. Robinson, Eds. Cambridge, UK: Springer, 2004, pp. 257-266.

Takeuchi et al. (1994). Simulated facial displays toward social interactions. *Proceedings of CHI94*, ACM Press.

VICTEC, (2003). VICTEC Project website www.victec.org, 2003.

Woods S., K Dautenhahn, J Schulz, (2004). The Design Space of Robots: Investigating Children’s Views. *RO-MAN Conference, 13th IEEE Intl. Workshop on Robot and Human Interaction and Communication. Kurashiki, Japan. Sept 2004.*

How May I Serve You? A Robot Companion Approaching a Seated Person in a Helping Context

K. Dautenhahn, M. Walters,
S. Woods, K. L. Koay, C. L. Nehaniv,
University of Hertfordshire,
College Lane,
Hatfield, UK, AL10 9AB
+44 (0) 1707 285113
K.Dautenhahn@herts.ac.uk

E. A. Sisbot, R. Alami,
T.Siméon
LAAS-CNRS
7, Avenue de Colonel Roche,
31077 Toulouse, FRANCE
+33 (0) 5 61 33 63 46
Emrah.Akin.Sisbot@laas.fr

ABSTRACT

This paper presents the combined results of two studies that investigated how a robot should best approach and place itself relative to a seated human subject. Two live Human Robot Interaction (HRI) trials were performed involving a robot fetching an object that the human had requested, using different approach directions. Results of the trials indicated that most subjects disliked a frontal approach, except for a small minority of females, and most subjects preferred to be approached from either the left or right side, with a small overall preference for a right approach by the robot. Handedness and occupation were not related to these preferences. We discuss the results of the user studies in the context of developing a path planning system for a mobile robot.

Categories and Subject Descriptors

A.m [**Miscellaneous**]: Human Robot Interaction – *Social Robots*

I.2.9 [**Artificial Intelligence**]: Robotics – *Mobile robots*

General Terms

Human Factors,

Keywords

Human-robot interaction, social robot, social spaces, personal spaces, user trials, live interactions

1. INTRODUCTION

If robots are to be used in office and domestic environments, they will have to encounter and interact with people. They must survive and carry out tasks in a disordered and unpredictable environment, safely and effectively. This paper presents the results from Human Robot Interaction (HRI) trials carried out at the University of Hertfordshire (UH). These results have then been used to inform and guide work carried out at the Laboratory for Analysis and Architecture of Systems at the Centre National de la Recherche Scientifique (LAAS-CNRS), to develop a task planning, motion planning, and control system that incorporates human social factors and preferences.

The work presented in this paper contributes to the COGNIRON Project [2005]. Part of this research into a cognitive robot companion investigates socially interactive robots [7] from a human-centred perspective, i.e. how robots could be useful in domestic environments; in particular the roles, tasks, and social behaviour(s) that will be necessary for robots to exhibit in order to integrate into everyday domestic situations. In order to study human-robot relationships, HRI trials using carefully devised test scenarios are conducted [18], where human responses and opinions can be collected using a variety of methods. A number of previous live HRI trials with human scaled PeopleBot^{TM} robots have been carried out [6, 17, 19, 20]. Other researchers have also investigated similar HRI trials with human sized robots including Dario et al. [4], Severinson-Eklundh et al. [16], Kanda et al. [9] and Hinds et al. [8].

Once the desired behaviour(s) for sociable robots capable of competent human-robot interactions are known, the challenge is then to incorporate the results into mobile robot path planning algorithms and control systems. At LAAS-CNRS progress has been made towards a motion planning framework that will allow the implementation of key criteria and parameters that can incorporate these results into the control system of a mobile robot that can be applied to human-centred environments. The presence of humans raises new issues for motion planning and control since the human's safety and comfort must be taken into account. The claim here is that a human-aware motion planner must not only consider safe robot paths, but also plan good, socially acceptable and legible paths.

There are a number of contributions in the literature where humans and robots co-exist in the same environment. These studies have frequently focussed only on the safety of the human [2, 10, 11, 21] and have failed to take *human comfort* into account. The planner presented here explicitly takes into account the human partners' safety and comfort by reasoning about accessibility, visual field, posture, gaze direction, relative distance to the robot and potential shared motions. Although several authors have proposed motion planning or reactive schemes with a consideration for humans, there is no contribution that has tackled this whole problem.

HRI'06. March 2-4, 2006, Salt Lake City, Utah, USA

2. The Live HRI Trials

This section presents results from two live HRI trials. First, a human-robot interaction *demonstration trial* event, which was run as part of an informal evening event at the AISB'05 Convention held at University of Hertfordshire in April 2005, and secondly, *follow-up trials* carried out in a controlled laboratory set-up, to re-test the results gained from the demonstration trial.

2.1 The HRI Trial Method

The trials were both carried out in converted seminar rooms where the scenario involved a robot using three different approach directions (front, left and right) to bring a seated subject an object (a TV remote control). The main aim of both trials was to establish subjects' preferences for the different robot approach directions. The demonstration event was conducted as part of an evening of entertainment for convention delegates, and involved different robot demonstrations. Spectators were present during the trials which were performed under non-laboratory conditions using 38 volunteers from the convention. The follow up study was carried out under controlled conditions with 15 subjects, and one of the main aims of this trial was to re-test the results obtained from the informal study.

2.1.1 The Trial Areas

The trial set-up was virtually identical for both trials and resembled a simulated living room with a chair and two tables. The subject was seated in the chair, which was positioned halfway along the rear wall (point (9), Fig.1), throughout the trial. To the left front, and right front of the chair, two tables were arranged (with room for the robot to pass by) in front of the chair. One of the tables had a television placed upon it; the other had a CD Radio unit. The robot was driven under direct remote control to the appropriate start position by an operator, but the robot's approaches to the subject were fully autonomous. The operator was seated at a table in the far corner of the room. Subjects were told that the robot would be controlled by the operator while it was driven to the three start positions, but would be approaching them autonomously to bring them the TV remote control. This was reinforced as the operator made notes and did not press any of the robot control keys (on the robot control laptop) while it approached the subject (Figure 1). The robot carried the remote control in a small basket suspended between the fingers of the lifting gripper. The remote control was placed in the basket prior to each experimental run. For each approach trial, the subject took the remote from the basket then replaced it ready for the next approach.

2.1.2 The HRI Trial Scenario

The same scenario was used for both HRI trials, introduced by the experiment supervisor. The context explained to the subjects was as follows: the subject had arrived home, tired after a long day at work and rested in an armchair (point (9), Fig.1). After looking around for the TV remote control, the subject then asked the robot to fetch it for them as they were too tired to get up. The robot then brought the remote control to the subject. It was explained to the subject that the robot was new to the household and it was necessary to find out which approach direction the subject preferred; either from the front (2), the left (1) or the right (3). The three possible paths taken by the robot are shown in Fig. 1. In order to justify the scenario of the robot fetching the remote control, one of the tables had a (switched off) TV set upon it. The other table had a CD-Radio unit. Our expectations prior to the trials were that subjects would prefer the approach from the front, since the robot was then fully visible at all times. Since many subjects, in particular in the demonstration trial, had never seen the robot before we assumed that they would feel most secure, comfortable, and 'in control' when the robot was fully visible so that its behaviour could be monitored easily.

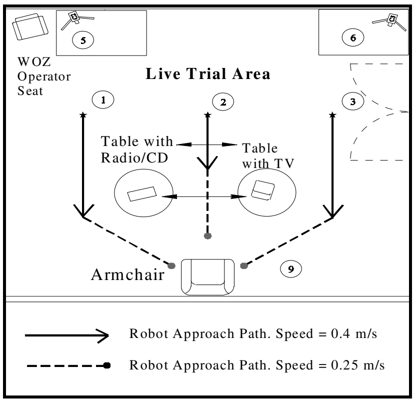


Figure 1. Live Trial Area

Figure 2. Examples of the demonstration HRI trial

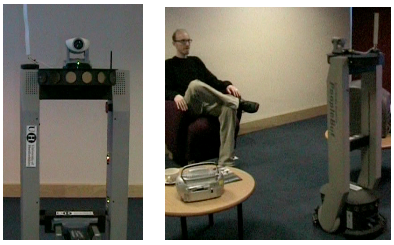


Figure 3. Example of the follow-up HRI trial.

2.1.3 Experimental Conditions

We were aware that the TV might be a natural focus of subjects' attention and may have influenced the choice of preferred robot approach direction. Therefore, for the controlled lab condition, half the trials were carried out with the TV on the left hand table, and the other half with the TV on the right hand table. Each subject experienced the robot approaching from three directions: front, left and right. To avoid any order effects, a counterbalanced order sequence covering all six possible permutations of the three robot approach directions was used. For the demonstration event, subjects experienced each approach direction only once, and for the controlled follow-up trials, each subject experienced the three robot approach directions twice, in a counterbalanced order.

2.1.4 Subject Sample Sets:

For the demonstration trial, 21 males (54%) and 18 females (46%) participated. The mean age of subjects was 36 years (range: 22-58). Thirty five subjects (95%) were right handed, and 2 subjects (5%) were left handed. All were delegates at the AISB'05 Convention. Fifteen subjects (9 (60%) males; 6 (40%) females) participated in the follow-up study. The mean age of this sample was 33 years (range 21-56 yrs). Only one subject was left handed. Four subjects were secretarial staff, 5 subjects were MSc students studying 'Artificial Intelligence', and the remaining 6 were research staff in the Computer Science Department at University of Hertfordshire. No subjects had previous exposure to the robots used in the trial. In the demonstration trial, some subjects had not sat straight in the chair (see Fig. 2). In the follow-up study subjects were made to sit straight with their feet to the front of the chair.

2.1.5 Procedure

For both trials, subjects completed a short introductory questionnaire to gain the necessary consent, and demographic details. At the end of each trial a semi-structured questionnaire was used to assess subject attitudes and preferences for the different robot approach directions and approach speed, as well as practicality issues. The questionnaires used for the follow-up trials were more extensive and included questions about the robot stopping distances, comfort levels and practicality for the different approach directions, rated according to a 5-point Likert scale. Subjects also participated in a semi-structured interview after the follow-up trial. The interview was carefully designed to eliminate leading questions. The main purpose of the structured interview was to assess subjects' views about the trial procedures and methodology, and find out how the trial could be improved from the participants' point of view. The subjects' reactions to both HRI trials were recorded by a single tripod mounted camera placed at an appropriate point at either (5) or (6) in Fig. 1.

2.2 Demonstration Trial Results

2.2.1 Overall Approach Direction Preferences:

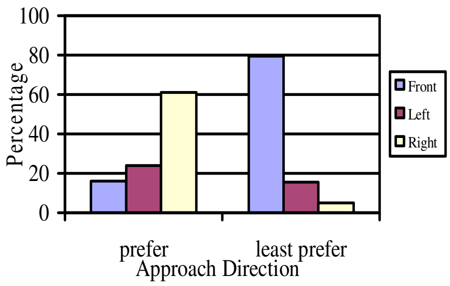


Figure 4. Demonstration trial: Robot to human approach direction preferences.

Figure 4 illustrates that 60% (N: 23) of subjects stated that they preferred the right robot approach direction, followed by 24% (N: 9) preferring the left approach and just 16% (N: 6) preferring the front approach. An overriding majority of subjects stated that they least preferred the frontal robot approach direction (N: 31, 80%). Very few subjects least preferred the left and right approach directions.

2.2.2 Gender Differences & Approach Direction Preferences

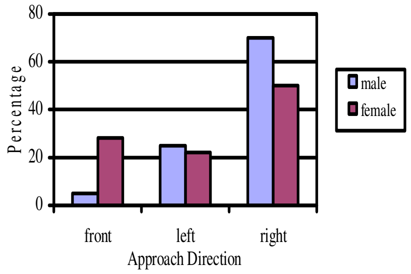


Figure 5. Male and female approach direction preferences

Chi-square cross-tabulations revealed a significant trend between gender and the preferred robot approach direction (X^2 (2, 38) = 3.77, p = 0.1). More females stated that they preferred the front robot approach direction compared to males, and more males preferred the right robot approach direction compared to females (see Figure 5). A significant relationship was found between gender and least preferred robot approach direction (X^2 (2, 39) = 7.09, p = 0.03). Significantly more males stated that they least preferred the front robot approach direction compared to females (males: 95%, females: 61%). More females stated that they least preferred the right robot approach direction compared to males (males: 0%, females: 11%).

2.2.3 Age, Handedness, and Approach Direction Preferences

Chi-square cross-tabulations revealed no significant relationships between age, handedness and approach directions preferred and least preferred.

2.2.4 Approach Distance

76% (N: 28) of subjects stated that the distance between them and the robot (0.5m ±0.1m) was 'about right', followed by 19% (N: 7) who felt that the robot was to 'too far' from them. Only 5% (N: 2) of subjects stated that the robot approached them too closely.

2.2.5 Practicality of Approach Directions

In addition to subjects rating which robot approach direction they preferred, ratings were given for how 'practical' they thought each approach direction was for the given task of delivering a TV remote control, according to a 5-point Likert scale (1 = not practical at all to 5 = very practical). A Friedman test for ordinal data illustrated that the rankings for approach direction practicality were significantly different from each other (X^2 (39, 2) = 12.11, p < 0.01). The mean rankings indicated that the front approach direction (mean ranking = 1.63) was rated as the least practical, and the right approach the most practical (mean ranking = 2.33), followed by the left (mean ranking = 2.04) approach direction.

2.2.6 *Comfort Ratings of Approach Directions*

Subjects were asked to rate how comfortable they felt with the different robot approach directions trials according to a 5-point Likert scale (1 = very uncomfortable, 5 = very comfortable). A Friedman test showed that the comfort level rankings for approach directions were significantly different from each other (X^2 (39, 2) = 29.38, $p < 0.001$). The mean rankings highlighted that subjects were the least comfortable with the front (mean ranking = 1.37) robot approach direction, and the most comfortable with the right approach direction (mean ranking = 2.49), followed by the left (mean ranking = 2.14).

2.3 Follow-Up Trial Results

2.3.1 *Approach directions most and least preferred*

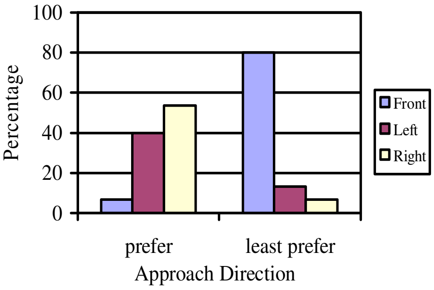


Figure 6. Follow-up trial: Least preferred and most preferred robot to human approach directions.

Results of the follow-up approach direction robot trials under laboratory conditions clearly demonstrated that the least preferred approach direction was the front approach. The right approach direction was the most preferred. These results are highly consistent, with the demonstration trial results (Figure 6).

2.3.2 *Robot Distance from the Subject*

For the robot's front approach direction stopping distance, 53% (N = 8) of subjects rated that the robot's stopping distance was too close. 27% (N = 4) of subjects rated that the robot's stopping distance was about right, and 20% rated that robot's stopping distance was too far. These results seem to indicate that a near majority of subjects rated that the front approach stopping distance was too close. In the case of subjects who rated the stopping distance as being too far for the front approach, we observed that these subjects usually had their legs stretched out in front of them causing the robot to stop when it reached the subject's feet rather than their arm for them to reach the TV remote control (due to the robot's stopping safety mechanism which had to be operational due to safety considerations). During the robot's approach from the left direction, 80% (N = 12) stated that the stopping distance was about right and 20% (N = 3) rated the stopping distance as being too far. During the robot's approach from the right of the subject 60% (N = 9) of subjects rated the stopping distance as about right, and 40% (N = 6) rated it as too far. It is interesting to note that no subjects thought the robot approached too closely from either left or right approach directions.

2.3.3 *Robot's Speed during the Trial*

The robots final approach speed to the subject was approximately 0.4 to 0.25 m/s, but was not finely controlled due to the inbuilt safety speed limiting mechanism. When subjects were asked to rate the robot's approach speed, 60% (N: 7) of participants rated that the speed was about right, and 40% (N: 6) of subjects rated that the robot's speed was too slow. None of the subjects rated that the robot's speed was too fast during the trials.

2.3.4 *Practicality and Comfort of the different Robot Approach Directions*

The front approach direction received the lowest practicality ratings for both the live and video trials. The right approach direction received the highest ratings of practicality followed by the left approach. The lowest mean comfort levels were found for the front robot approach direction. The highest comfort level rating was found for the right approach direction followed by the left approach direction. No significant differences were found between most preferred approach direction and least preferred approach direction for gender, subject handedness (whether subject was left or right handed), and occupation.

2.4 Combined Results of Demonstration & Follow-Up Trials

In light of the comparable HRI trial methodologies and the high degree of agreement between the results from the informal demonstration trials and formal follow-up trials, the results from both trials were combined to form one dataset from the 55 subjects who participated in both trials. Thirty males (56%) and 24 females (44%) in total participated in the robot approach direction trials. The mean age of subjects was 36 years (range: 21-58, SD: 11.54). Forty nine subjects (94%) were right handed, and 3 subjects (6%) were left handed.

2.4.1 *Trial Preferences:*

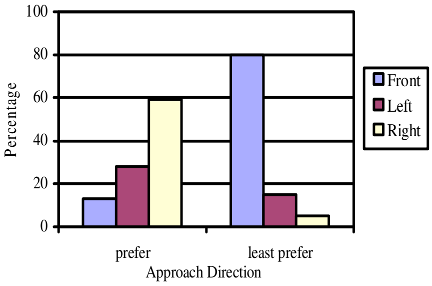


Figure 7. Combined trial results: Overall robot to human approach direction preferences

Figure 7 illustrates that 59% (N: 31) of subjects stated preferring the right robot approach direction, followed by 28% (N: 15) who preferred the left approach, and just 13% (N: 7) preferred the front approach. An overriding majority of subjects stated least preferring the front robot approach direction (N: 43, 80%). Few subjects least preferred the left and right approach directions.

2.4.2 *Practicality of Approach Directions*

A Friedman test for ordinal data illustrated that the rankings for approach direction practicality were significantly different from each other (X^2 (54, 2) = 21.87, $p < 0.001$). The mean rankings indicate that the front approach direction (mean ranking = 1.55) was rated as the least practical, and that the right approach was the most practical (mean ranking = 2.34), followed by the left (mean ranking = 2.11) approach direction.

2.4.3 Comfort Ratings of the Approach Directions

Results from a Friedman test showed that the comfort level rankings for approach directions were significantly different from each other (X^2 (54, 2) = 47.78, $p < 0.001$). The mean rankings highlight that subjects were the least comfortable with the front (mean ranking = 2.43) robot approach direction, and the most comfortable with the right approach direction (mean ranking = 4.15), followed by the left (mean ranking = 3.76).

2.4.4 Gender Differences

Chi-square cross-tabulations revealed a significant association between gender and the robot approach direction preferred (X^2 (2, 53) = 5.83, $p = 0.05$). More females stated preferring the robot front approach direction compared to males, and more males preferred the right robot approach direction compared to females (See Figure 8). A small significant relationship was found between gender and least preferred robot approach direction (X^2 (2, 54) = 5.72, $p = 0.06$). More males stated least preferring the front robot approach direction compared to females (males: 90%, females: 67%). More females stated least preferring the left (males: 10%, females: 21%) and right robot approach direction compared to males (males: 0%, females: 13%). Independent measures t-tests revealed a trend for males (M = 4.37) to rate the right robot approach direction as more comfortable compared to females (M = 3.88) (t (52) = 1.74, $p = 0.08$). No further significant gender differences were revealed for comfort ratings of the front and left robot approach directions.

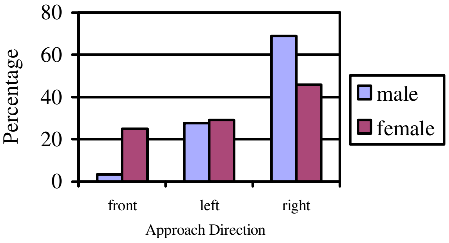


Figure 8. Combined results: Male and female preferences

Independent measures t-tests were calculated to examine gender differences and ratings of the practicality of the robot approach directions. Significant differences were found for the practicality of the front approach direction (t (52) = -2.46, $p = 0.02$). Females rated the front approach direction as significantly more practical compared to males (males M = 2.60, females M = 3.38). No further significant differences were found between gender and practicality ratings for the left and right approach directions.

2.4.5 Age, Handedness, and Approach Direction Preferences

Chi-square cross-tabulations revealed no significant relationships between age, handedness, approach directions most and least preferred, comfort ratings of the approach directions, and practicality ratings of the approach directions.

2.4.6 Comments made by Subjects about the Three Robot Approach Directions.

Subjects were asked to provide details about the reasons for preferring and least preferring particular robot approach directions. The most frequently cited comments are provided in tables 1 and 2.[1]

Table 1. Reasons why subjects preferred a particular approach direction.

Preferred Front Approach Direction
Front approach direction was easy to reach for the TV remote control
The effort needed to reach for the remote control was the least, but this was still not close enough
Preferred Left Approach Direction
I felt the most relaxed and comfortable during this approach
Preferred this approach as I am left handed
This approach was the quickest and most direct
This approach felt the most natural
It was the most convenient for the robot to approach this way.
Preferred Right Approach Direction
I felt the most comfortable with this approach
This approach seemed the most natural
I am right handed, so it was the easiest way to take the remote control
This approach seemed to be the quickest
This approach because it was always in my field of vision

Table 2. Reasons why subjects least preferred a particular approach direction.

Least Preferred Front Approach Direction
I had to move forward to reach for the remote control, the robot was too far away from me
This approach was slightly threatening
This approach was just a little bit too close for comfort
Seemed too aggressive
The robot was always looking a me
I was concerned about the robot running into me during this approach
This approach was intimidating
Least Preferred Left Approach Direction
Didn't like left approach as I am right handed
It was difficult for me to reach for the remote control
I felt awkward reaching across with me left hand
It felt like I had to reach further for the left approach
The robot was not in my line of vision during the left approach
Least Preferred Right Approach Direction
Least preferred this approach because I am left handed
The robot felt like it was behind my back during this approach

Implications of User Studies for Robot Motion Planning

Today, classical motion planning methods [12] are quite efficient at locating feasible paths. However, the presence of humans in the environment drastically changes the notion of acceptable paths. In a human-robot interaction context, the computed paths do not only need to be collision-free but must also take into account human

---

[1] Due to space limitations, only the most frequently cited comments are shown.

comfort. This is illustrated in figure 9, which shows two paths produced by a classical motion planner. Both paths are inconvenient since one path passes too close to the wall, causing the human to be surprised, and the other passes behind the human resulting in discomfort. The HRI studies reported in the previous section, and others [1, 13, and 19] highlight a number of properties that must be taken into account when dealing with humans. Only limited studies have considered comfort and legibility issues, often in an ad hoc manner. A new technique is described that integrates additional constraints in a more generic way. In these steps of our work, we assume that the final positions of the paths are already calculated.

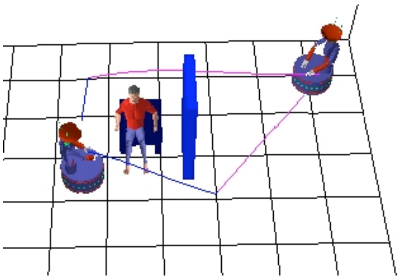


Figure 9. Two paths found by classical motion planning systems

We introduce three criteria to the motion planning stage to ensure safety and comfort. The robot must take into account these three criteria at the planning stage along with the more common aspects of path planning such as obstacle avoidance. Each criterion is represented by human-centred costs stored in a 2D grid:

Safety Criterion: This focuses on ensuring safety by controlling the distance between the robot and human. The robot, if possible, must avoid approaching the human too closely, and in some cases (i.e. no physical interaction) the robot must not be able to pass through a certain perimeter around the human. However, the robot must be able to approach the human to allow interactions to occur (for example to pass an object to a human). Hence, this distance between the robot and the human is not uniform and fixed, but depends on the type of human-robot interaction, in addition to the human preferences, and physical abilities. For instance, the user studies presented above are reflected by a configuration of costs that favours approach motions by the side (Fig 10).

Visibility Criterion: Human comfort is a key issue when dealing with HRI scenarios, and some properties can be extracted from this issue. In particular, humans generally feel more comfortable when the robot is within their field of vision. Therefore, a "visibility criterion", is used to help the robot to stay, during its motions, in the human's field of view. The visibility grid is constructed according to costs reflecting the effort required by the human to get the robot in his field of view. Grid points located in a direction for which the human has only to move his eyes have a lower cost than positions requiring head turning in order to get the robot in the field of view. Also, when the robot is far away from the human, the effect of visibility must decrease, and beyond a certain distance it must be negligible.

Hidden Zones: In the grids presented above, the costs are calculated without taking into account obstacles in the environment. However, obstacles in close vicinity to the human can have various effects on safety and visibility issues. If the robot is behind an obstacle, the human might feel comfortable because the obstacle would block the direct path between the human and the robot. Therefore, the safety criterion must be cancelled in zones located behind the obstacles. In contrast, as the robot passes behind an obstacle and becomes hidden, and the human cannot see the robot, the visibility costs no longer correspond to physical realities. To handle this issue, we introduce a further criterion termed, "hidden zones criterion". This criterion helps to determine better costs for positions hidden from the human by obstacles. An important effect of obstacles for human comfort is the "surprise factor". When the robot is hidden by an obstacle close to the human, and suddenly appears in the human field of vision, it can cause surprise and possibly fear. To avoid this effect, we must discourage the robot to pass behind an obstacle too closely, and must allow it to get into the human's field of view when sufficiently far from the human. This can be done by adding costs to the zones hidden from the subject's view by the obstacles. The costs in the hidden zone grid are inversely proportional to the distance between the human and the robot so that the robot chooses to keep a distance from back sides of the obstacles that are close to humans.

Once the safety, visibility, and hidden zones grids have been computed (Fig. 10), they are merged into one single grid where the robot will search for a minimum cost path.

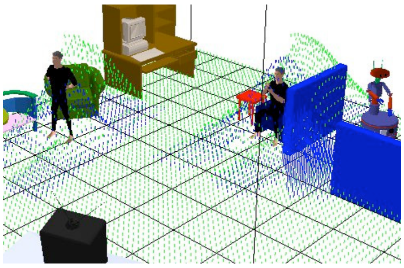


Figure 10. The "safety", the "visibility" and the hidden zones" grids. The height of a point corresponds to the cost of that point. The grids were modified to correspond to the results of the user studies.

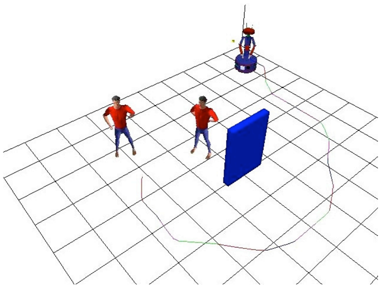


Figure 11. A human friendly path calculated automatically by the planner. Note the robot does not choose the shortest path and prefers a path that avoids it "to burst" near the human.

Different ways, depending on the task and on the balance between criteria, can be used to aggregate the grid costs.

For example, for an urgent task, the importance of the visibility grid is less than the safety grid so that the robot does not take visibility largely into account. Once the final grid is computed, the cells corresponding to the obstacles in the environment are labelled as forbidden and an A* search is performed to find minimum-cost path between two given positions of the robot. Since only crossing the obstacles and humans are forbidden, with this algorithm we guarantee to find a path if it exists.

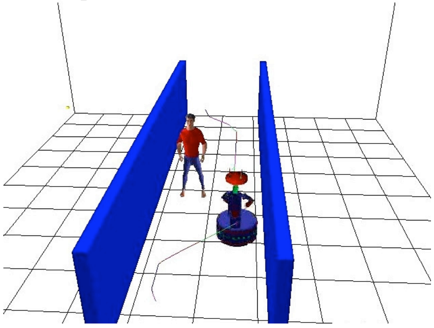


Figure 12. A Hallway scenario. The planner automatically plans a trajectory that allows the robot to pass next to the human without causing any discomfort. Note that as the robot does not immediately take a position behind the human, it avoids causing any discomfort when it is invisible to him.

The computed paths shown in Figures 11 and 12 are collision-free and also take into account the human's comfort and safety.

3. Conclusions

Results from the two HRI trials indicate that a large majority of human subjects, when seated, preferred a robot to approach from either the left or right side. The frontal approach was seen as uncomfortable, impractical, in some cases even threatening or confrontational, and should thus be avoided. This result is in line with human-human situations where standing or sitting at an angle of 45 degrees to each other can reduce feelings of aggression and confrontation [14]. However, the side that an individual human will actually prefer, left or right, depends to a large extent on the preferences of the individual concerned. The results do show that there is a bias towards the right hand side. This may be related to the fact that most of the trial subjects were right handed (in common with most of the population in general). Therefore, for a robot which is bringing an object to a seated human whose preferences are not known, it should always avoid a frontal approach and if (physically) convenient and consistent with the particular task then approach from the right. If the seated humans' approach direction preferences are known, then the robot should approach from the preferred direction whenever convenient[2]. It should be noted here that a human subject will not be unduly disturbed if their approach preference with regard to which side are not followed.

There were some perceived gender differences with regard to approach direction, with some females actually preferring a frontal approach direction, whereas slightly more males than females preferred a right side approach over other directions. From psychological studies [14] it has been found that women tend to stand slightly closer to one another, face each other more, and touch each other more, compared to men interacting with other men. That women tend to face each other more could possibly account for the fact that women in our studies more frequently preferred the robot to approach from the frontal direction compared to men, although this issue needs further investigation.

In the follow-up trials, no subjects thought that the robot came too close from either the right or left side directions, though a majority thought the front approach distance was too close. In all cases the robot approached to no closer than 50cm, which was the inbuilt safety collision avoidance distance of the robot. It has been noted that there are cultural differences in personal spatial zones [14][3]. However, although some subjects in the HRI trials may have originated from other countries and cultures, all the subjects had been resident in the UK and therefore could be presumed to adopt human-human social distances similar to those of the average UK population. Therefore, regional, cultural or ethnic origin information was not asked (or controlled) for in the studies[4].

Most subjects stated that the robot moved too slowly or about right at 0.4m/s, while nobody rated that the robot moved too fast. This suggests that (especially after a longer habituation period), most subjects would prefer the robot to move at a faster speed. It would therefore be reasonable to set the default robot speed at a relatively slow 0.4m/s and then perhaps increase the approach speed over time or in response to the user's wishes or preferences.

The robot used in the trials only had a simple short reach gripper, so the object was presented to the subject in a simple lifting tray. If a longer manipulator or arm was fitted, the results obtained may well be very different. It is desirable to perform further trials with various robots fitted with various types of arms or manipulators to see what effect they may have on user preferences. Also, long term trials are needed to investigate the effect on people of longer periods of exposure to robots. It would also be interesting to perform human-human studies to complement the work presented here. However, the primary focus of this paper is on robot to human approach direction preferences.

The human-aware motion planner is in its first steps of development and implementation. It requires further experiments to customize and validate the planner for live HRI situations. We are planning to implement this motion planner along with task reasoning capabilities [3] into a real robot that must have sufficient

---

[2] Deriving such 'social rules' for robots from empirical HRI studies is part of an attempt to develop a *robotic etiquette*, cf. B. Ogden, K. Dautenhahn (2000) Robotic Etiquette: Structured Interaction in Humans and Robots, in Proc SIRS2000, 8th Symposium on Intelligent Robotic Systems, The University of Reading, England, 18-20 July 2000.

[3] For example, many southern Europeans and Japanese have an intimate distance (reserved for close friends and family) of only 20-30cm compared to 46-122cm of the Americans and northern Europeans. Europeans might refer to Asians as 'pushy' and 'familiar' and Asians might refer to Europeans and Americans as 'cold' and 'stand-offish'. There are also differences in rural vs. urban spatial zones. People raised in more rural, less populated areas need more personal space, than those raised in densely populated cities.

[4] A specific study which investigates in more detail human robot approach distances using PeopleBot^{TM} robots is given in Walters et al. [20].

human perception capabilities such as determination and tracking of various features like human-body posture, head orientation, hand configuration and gaze direction. In the execution stage of the plan, the robot must be highly reactive to changes in the environment. Using path deformation approaches can ensure this reactivity.

Joint work as described in this paper will ultimately contribute to the development of *interaction-aware robots* [5], i.e. robots that are sensitive to the social context they are embedded in. This is a vital requirement for all those robotics applications where human contact and acceptability plays a vital part, as it is the case in domestic, healthcare and other applications. The challenge to develop robots that are not only ‘doing the right thing’, but ‘doing the thing right’ [15] can only be tackled in a interdisciplinary endeavour involving psychologist as well as roboticists and HRI experts.

4. ACKNOWLEDGMENTS

The work described in this paper was conducted within the EU Integrated Project COGNIRON ("The Cognitive Robot Companion") and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020.

5. REFERENCES

[1] Althaus, P., Ishiguro, H., Kanda, T., Miyashita, T. and Christensen, H.I., Navigation for Human-Robot Interaction Tasks. in Proc. IEEE Int. Conf. on Robotics & Automation, (New Orleans, USA, 2004).

[2] Bicci, A. and Tonietti, G. Fast and soft arm tactics: Dealing with the safety-performance trade-off in robot arms design and control. IEEE Robotics and Automation Magazine, 11 (2), 12-21.

[3] Clodic, A., Montreuil, V., Alami, R. and Chatila, R., A decisional framework for autonomous robots. in Proc. 14th IEEE Int. Workshop on Robot & Human Communication (RO-MAN), (Nashville, USA, 2005), 543-548.

[4] Dario, P., Guglieimelli, E. and Laschi, C. Humanoids and personal robots: Design and experiments. Journal of Robotic Systems, 18 (12), 673-690.

[5] Dautenhahn, K., Ogden, B. and Quick, T. From embodied to socially embedded agents - Implications for interaction-aware robots. Cognitive Systems Research, 3 (3), 397-428.

[6] Dautenhahn, K., Woods, S., Kaouri, C., Walters, M., Koay, K.L. and Werry, I., What is a robot companion - Friend, assistant or butler? in Proc. IEEE IROS, (Edmonton, Canada, 2005), 1488-1493.

[7] Fong, T., Nourbakhsh, I. and Dautenhahn, K. A survey of socially interactive robots. Robotics and Autonomous Systems, 42, 143-166.

[8] Hinds, P., Roberts, T. and Jones, L. Whose job is it anyway? A study of human-robot interaction in a collaborative task. Human Computer Interaction, 19, 151-181.

[9] Kanda, T., Hirano, T., Eaton, D. and Ishiguro, H. Interactive robots as social partners and peer tutors for children: A field trial. Human Computer Interaction, 19 (1-2), 61-24.

[10] Kulic, D. and Croft, E., Safe Planning for Human-Robot Interaction. in Proc. IEEE Int. Conf on Robotics & Automation, (New Orleans, USA, 2004).

[11] Kulic, D. and Croft, E., Strategies for safety in human-robot interaction. in Proc. IEEE Int. Conf. on Advanced Robotics, (2003), 810-815.

[12] Latombe, J.C. Robot motion planning. Kluwer Academic Publishers, Boston, USA, 1991.

[13] Pacchierotti, E., Christensen, H.I. and Jensfelt, P., Human-robot embodied interaction in hallway settings: A pilot user study. in Proc. 14th IEEE Int. Workshop on Robot & Human Interactive Communication (RO-MAN), (Nashville, USA, 2005), 164-171.

[14] Pease A. and Pease B. The definitive book of body language. London. Orion Books Ltd, 2004.

[15] Sengers, P., Do the right thing: An architecture for action expression. in Proc. Second Int. Conf. on Autonomous Agents, (1998), 24-31.

[16] Severinson-Eklundh, K., Green, A. and Hüttenrauch, H. Social and collaborative aspects of interaction with a service robot. Robotics and Autonomous Systems, 42. 223-234.

[17] Te Boekhorst, R., Walters, M.L., Koay, K.L., Dautenhahn, K. and Nehaniv, C. A study of a single robot interacting with groups of children in a rotation game scenario. In Proc. of IEEE CIRA 2005, (Finland, 2005).

[18] Walters, M., Woods, S., Koay, K.L. and Dautenhahn, K., Practical and methodological challenges in designing and conducting interaction studies with human subjects. In Proc. of AISB'05, (University of Hertfordshire, Hatfield, UK, 2005), 110-119.

[19] Walters, M.L, Dautenhahn, K., Te Boekhorst, R., Koay, K.L, Kaouri, C, Woods, S, Nehaniv, C, Lee D. and Werry, I. The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment. .in Proc. 14th IEEE Int. Workshop on Robot & Human Communication (RO-MAN), (Nashville, USA, 2005), 347-352.

[20] Walters M. L, Dautenhahn K, Koay K. L, Kaouri C, te Boekhorst R, Nehaniv C. L , Werry I and Lee D. Close encounters: Spatial distances between people and a robot of mechanistic appearance. Proc. IEEE-RAS Humanoids 2005, December 5-7, 2005, Tsukuba, Japan., 450-455

[21] Zim, M., Khatib, B., Roth, B. and Salisbury, J.K. Playing it safe (human friendly robots). IEEE Robotics and Automation Magazine, 11 (2), 12-21.

ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Robotics and Autonomous Systems ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Robotics and Autonomous Systems

www.elsevier.com/locate/robot

Safe proactive plans and their execution

K. Madhava Krishna[a], R. Alami[b], T. Simeon[b,*]

[a] International Institute of Information Technology, Hyderabad, India
[b] LAAS-CNRS, 31077 Toulouse Cedex, France

Received 29 May 2004; received in revised form 28 October 2005; accepted 29 October 2005

Abstract

We present in this paper a methodology for computing the maximum velocity profile over a trajectory planned for a mobile robot. Environment and robot dynamics as well as the constraints of the robot sensors determine the profile. The planned profile is indicative of maximum speeds that can be possessed by the robot along its path without colliding with any of the mobile objects that could intercept its future trajectory. The mobile objects could be arbitrary in number and the only information available regarding them is their maximum possible velocity. The velocity profile also enables one to deform planned trajectories for better trajectory time. The methodology has been adopted for holonomic and non-holonomic motion planners. An extension of the approach to an online real-time scheme that modifies and adapts the path as well as velocities to changes in the environment such that both safety and execution time are not compromised is also presented for the holonomic case. Simulation and experimental results demonstrate the efficacy of this methodology.

Keywords: Robot motion planning and control; Safe robot motions; Environmental constraints; Sensory constraints

1. Introduction

Several strategies exist for planning collision-free paths in an environment whose model is known [9]. However, during execution, parameters such as robot and environment dynamics, and sensory capacities need to be incorporated for safe navigation. This is especially so if the robot navigates in an area where there are other mobile objects such as humans. For example in Fig. 1, the robot would be required to slow down as it approaches the doorway, in anticipation of mobile objects emerging from there, even if it does not intend to make a turn through the doorway.

A possible means to tackle the above problem at the execution stage is to always navigate the robot at very low speeds. In fact, reactive schemes such as the nearness diagram approach [11] operate the robot at minimal velocities throughout the navigation. However, incorporating the computation of a velocity profile at the planning stage would circumvent not only the problem of conservative velocities throughout navigation but would also allow for a modification of the trajectory to achieve lower time (Fig. 2).

We present in this paper a novel proactive strategy that incorporates robot and environment dynamics as well as sensory constraints into a collision-free motion plan. By proactive we mean that the robot is always in a state of expectation regarding the possibility of a mobile object impinging onto its path from regions invisible to its sensor. This proactive state is reflected in the velocity profile of the robot, which guarantees that in the worst-case scenario, the robot will not collide with any of the moving objects that can interfere with its path. The ability of the algorithm to compute a priori velocities for the entire trajectory accounting for objects moving in arbitrary directions is the essential novelty of this effort.

As is always the case, planned paths and profiles need constant modification at the execution stage due to changes in the environment. For example a profile and path that was planned for an environment with a closed doorway needs to be modified during real time if the doorway is found open. Also addressed in this article is the problem portrayed in Fig. 3. Given an initial trajectory planned for a particular environment, how does the robot modify its trajectory while new objects (not necessarily intersecting the robot's trajectory) are introduced into the environment such that the basic philosophy of ensuring safety as well as reducing time lengths of the path continue to

---

[*] Corresponding author.
E-mail address: nic@laas.fr (T. Simeon).

0921-8890/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.robot.2005.10.008

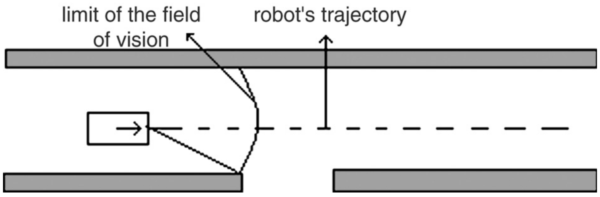


Fig. 1. A safe robot has to slow down while approaching the doorway.

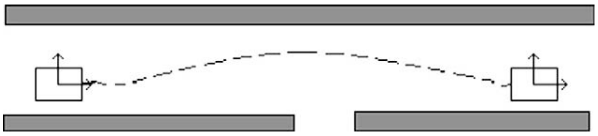


Fig. 2. A longer path can be faster due to higher speed.

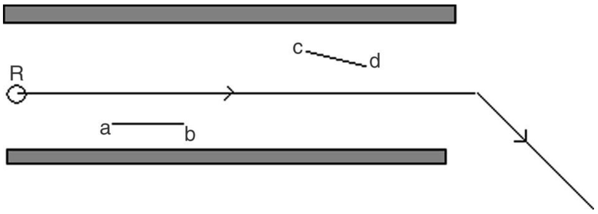


Fig. 3. How does the robot adapt its path in the presence of new segments (a, b) and (c, d) while maintaining safe velocities?

be respected? Simulation and experimental results are presented to indicate the efficacy of the scheme. In [1] we reported how the maximum velocity profiles can be computed for any generic planner and in [8] we presented initial simulation and experimental results of the reactive version of [1].

Related work can be cited in the areas of modifying global plans using sensory data obtained during execution for overcoming uncertainty accumulated during motions [3] and those that try to bridge the gap between planning and uncertainty [10] or planning and control [7,2]. The velocity obstacle concept [13,5] bears resemblance to the current endeavour in that they involve selection of a robot velocity that avoids any number of moving objects. The difference is that in the present approach the only information about the mobile object available is the bound on velocity. The direction of motion and the actual velocities are not known during computation of the velocity profile. The work of Stachniss [14] also involves considering the robot's pose and velocities at the planning phase. A path is determined in the (x, y) space and a subgoal is chosen. A sequence of linear and angular velocities, (v, w), is furnished until the subgoal is reached. In [12] a policy search approach is presented that projects a low dimensional intermediate plan to a higher dimensional space where the orientation and velocity are included. As a result better motion plans are generated that enable better execution of the plan by the robot. The current effort has similarities to [12], at the planning level but also extends it to a suitable reactive level in the presence of new obstacles encountered during execution. Similarly the dynamic window approach [16] and the global dynamic window method of Brock et al. [17] both incorporate the dynamics and the kinematics of the robot for a reactive collision avoidance system. Incorporating the dynamics and searching in the space of velocities overcome the problems of purely geometric methods. However, these methods do not speak of modifying the path in order to reduce its time-length and the dynamics of the environment does not affect the computation of the velocity profile, which makes our approach different from those mentioned above.

2. Problem definition

The following problems are addressed in the paper, given:

- A robot $\mathcal{R}$ modelled as a disc and equipped with an omnidirectional sensor having a limited range R_{vis}. We call C_{vis} the visibility circle, centred at the robot's position with radius R_{vis}. The paths of $\mathcal{R}$ are sequences of straight segments or straight segments connected with circular arcs of radius ρ in case of a non-holonomic robot. The robot's motion is subject to dynamic constraints simply modelled by a bounded linear velocity $v \in [0, v_{rm}]$ and a bounded acceleration $a \in [-a_{-m}, a_m]$. The maximum possible deceleration a_{-m} need not equal the maximum acceleration a_m.
- A workspace cluttered by static polygonal obstacles $\mathcal{O}_i$. The static obstacles can hide possible mobile objects whose motions are not predictable; the only information is their bounded velocity v_{ob}.

Problem 1. Given a robot's path $\tau(s)$ computed by a standard planner [9], determine the maximal velocity profile $v\tau(s)$ such that, considering the constraints imposed by its dynamics, the robot can stop before collision occurs with any of the mobile objects that could emerge from regions not visible to the robot at position $s \in \tau(s)$. For example, the velocity profile dictates that the robot in Fig. 1 slow down near the doorway in expectation of mobile objects from the other side. We call $MP = (\tau(s), v_\tau(s))$ a **robust motion plan**. The velocity profile allows us to define the time $T(\tau)$ required for the robust execution of path τ:

$$T(\tau) = \int_0^L \frac{\mathrm{d}s}{v_\tau(s)}.$$

Problem 2. Modify the planned trajectory such that the overall trajectory time $T(\tau)$ is reduced. For example, the path of Fig. 2, albeit longer than the one of Fig. 1, is traversed in a shorter time.

Problem 3. Adapt the path and velocities reactively in the presence of new objects not a part of the original workspace such that the criteria of safe velocities and reduced time of path continue to be respected. This is illustrated in Fig. 3.

3. From path to robust motion plan

The procedure for computing the maximum velocity profile $v_\tau(s)$ delineated in Sections 3.1, 3.2 and 3.3 addresses the first problem. The constraints imposed by the environment on the

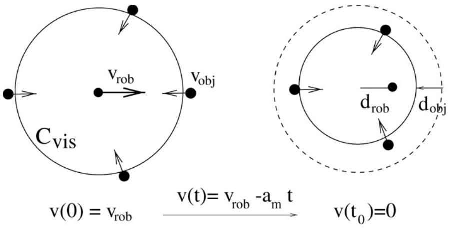


Fig. 4. Mobile objects may appear anywhere on C_{vis}'s contour.

robot's velocity are due to two categories of mobile objects. The first category consists of mobile objects that could appear from anywhere outside the boundary of the visibility circle C_{vis}. The second category involves mobile objects that could emerge from shadows created in C_{vis} due to stationary objects.

3.1. Velocity constraints due to the environment

No obstacles in C_{vis}

In the simple case where the robot's position is such that no static obstacle lies inside C_{vis}, a moving object may appear (at time $t = 0$) anywhere on C_{vis}'s boundary (Fig. 4). Let V_{rb} denote the maximum possible robot velocity due to a mobile object at the boundary. At time $t_0 = v_{rb}/a_{-m}$ (i.e., when the robot is stopped), the distance crossed by the object is $d_{obj}(v_{rb}) \le v_{ob}v_{rb}/a_{-m}$. Avoiding any potential collision imposes that $R_{\text{vis}} \ge d_{rb}(v_{rb}) + d_{obj}(v_{rb})$, where $d_{rb} = v_{rb}^2/2a_{-m}$. The condition relates v_{rb} to the sensor's range R_{vis} as:

$$v_{rb} = -v_{ob} + \sqrt{v_{ob}^2 + 2a_{-m}R_{\text{vis}}}. \tag{1}$$

Influence of shadowing corners

Static obstacles lying inside C_{vis} may create shadows (e.g., see the grey region of Fig. 5) which contain mobile objects. The worst-case situation occurs when the mobile object remains unseen until it arrives at the *shadowing corner* of a polygonal obstacle. Since the mobile object's motion direction is not known it is best modelled for a worst-case scenario as an expanding circular wave of radius $v_{ob}t$ centred at (d, θ)

$$(X(t) - d\cos\theta)^2 + (Y(t) - d\sin\theta)^2 = v_{ob}^2 t^2.$$

Let us first consider that the robot's path τ is a straight segment. Considering that the intersections between the circular wave and the robot's segment path should never reach the robot before it stops at time $t_0 = v_{rs}/a_{-m}$ yields the following velocity constraint:

$$v_{rsv}^4 - 4(a_{-m}d\cos\theta + v_{ob}^2)v_{rsv}^2 + 4a_{-m}^2 d^2 \ge 0. \tag{2}$$

Here v_{rsv} is the maximum possible robot velocity due to the shadowing vertex under consideration. The solution of Eq. (2) gives v_{rsv}, as a function of (d, θ).

This solution only exists under the condition $v_{ob} > \sqrt{a_{-m}d(1-\cos\theta)}$, i.e., when the object's velocity v_{ob} is sufficiently high to interfere with the robot's halting path. Otherwise, the shadowing corner does not constrain the robot's velocity, which can be set to v_{rm}, the maximum bound on robot's velocity.

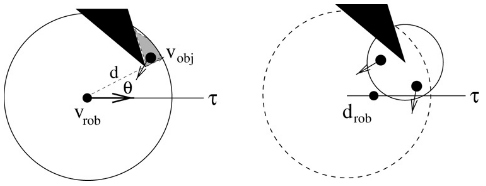


Fig. 5. Mobile objects may also appear from the shadows of static obstacles.

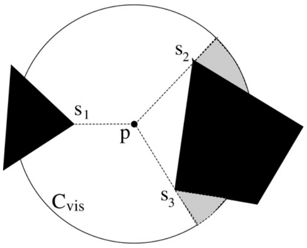


Fig. 6. Shadowing corners: among the three vertices of $\mathcal{V}(p)$, only s_2 and s_3 create shadows (the line going through s_1 is not tangent to the left obstacle).

Similar reasoning can be applied to the case where the robot traverses a circular arc path of radius ρ. This case, however, leads to a nonlinear equation that needs to be solved numerically to derive the maximal velocity [4]. The expression that needs to be solved for computing the maximum velocity at a given point on a circular arc is of the form:

$$((v_{rsv}^2 v_{ob}^2)/a_{-m}^2) + 2\rho^2\cos(v_{ob}^2/2a_{-m}\rho) + 2d\rho\sin((v_{ob}^2/2a_{-m}\rho) - \theta) = d^2 + 2\rho^2 - 2d\rho\sin\theta. \tag{3}$$

3.2. Computing the shadowing corners

The problem of determining the set of shadowing corners needed for the velocity computation in Section 3.1 is the problem of extracting those vertices of the polygonal obstacle to which a ray emitted from the robot's centre is tangential (Fig. 6). The set of shadowing corners can be easily extracted from an algorithm that outputs the visibility polygon [15] as a sorted list of vertices.

3.3. Computing the velocity profile $v_\tau(s)$

While the methodology for computing the maximum velocity profile delineated here is essentially for a holonomic path, its extension to the non-holonomic case is not very difficult.

1. A holonomic path τ, consisting of a sequence of straight line segments ab, bc, cd (Fig. 7), is deformed into a sequence of straight lines and clothoids to ensure continuity of velocities at the bends [6]. The maximum deviation from an endpoint to its clothoidal arc (depicted as e in Fig. 7) is dependent on the nearest distance to an object from the endpoint under consideration.

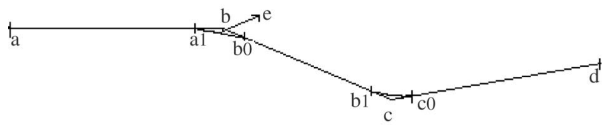


Fig. 7. A holonomic path deformed into a sequence of straight segments and clothoidal arcs.

2. The linear velocity along a clothoid is a constant and the maximum possible linear velocity considering robot dynamics alone is calculated for each of the clothoidal arc $a1b0, b1c0$ according to [6] and is represented as $v_c(a1)$, $v_c(b1)$.
3. The straight segment $aa1$ is discretized into M equally spaced points, excluding the endpoints of the segment, namely a and $a1$. We denote the first such point as a_1 and the last such point as a_M. The point of entry into the clothoid, namely $a1$, is also denoted as a_{M+1}.
4. For each of the N points, a_i, the steps 4a to 4e are repeated.

4a. The maximum possible velocity that a robot could have such that it can come to a halt before colliding with objects that enter into the robot's field of vision from the boundary is computed as $v_{rb}(a_i)$ according to Eq. (1).

4b. The velocity of the robot due to stationary obstacles inside the robot's field of vision that create shadows is computed as $v_{rsv}(a_i)$ according to Eq. (2). The minimum of all the velocities due to such vertices is found and denoted as $v_{rs}(a_i)$.

4c. The maximum possible velocity of the robot at a_i due to the environment is then computed as

$$v_{re}(a_i) = \min(v_{rb}(a_i), v_{rs}(a_i)). \tag{4}$$

4d. The velocity of the robot at a_i due to its own dynamics is given by

$$v_{rd}(a_i) = \sqrt{v_r^2(a_{i-1}) + 2a_m s(a_i, a_{i-1})}. \tag{5}$$

The above equation is computed if $v_{re}(a_i) > v_r(a_{i-1})$. Here $s(a_i, a_{i-1})$ represents the distance between the points a_i and a_{i-1}. a_m represents the maximum acceleration of the robot.

4e. The eventual velocity at a_i is given by

$$v_r(a_i) = \min(v_{rd}(a_i), v_{re}(a_i), v_{rm}). \tag{6}$$

Here v_{rm} represents the maximum robot velocity permissible due to servo motor constants.

5. The velocity at the endpoint $a1$ is computed as $v_r(a1) = \min(v_r(a1), v_c(a1))$ and this would be the linear velocity with which the robot would traverse the clothoid.
6. Steps 6a and 6b are performed by going backwards on each of the N points from a_N to a_1.

6a. If $v_r(a_i) > v_r(a_{i+1})$ then the modified maximum possible velocity at a_i is computed as

$$v_{rd}(a_i) = \sqrt{v_r^2(a_{i+1}) + 2a_{-m} s(a_i, a_{i+1})}. \tag{7}$$

6b. Finally, the maximum safe velocity at a_i is given as $v_r(a_i) = \min(v_r(a_i), v_{rd}(a_i))$.

7. Repeat steps 3 to 6 for all the remaining straight segments to obtain the maximal velocity profile over a given trajectory τ as $v_\tau(s) = \{v_r(a), v_r(a_1), \ldots, v_r(a_N), v_r(a1), v_r(b_1), \ldots, v_r(d)\}$.

3.4. Modifying the planned trajectory for lower time

The knowledge of the maximum velocity profile over a trajectory is utilized to tackle the problem posed in Section 2 of reducing the overall trajectory time of the path. The procedure for reducing the trajectory time at the planning stage involves random deformation of the planned path and evaluating time along this path. The modified path becomes the new trajectory if the time along it is less than that along the original trajectory. The process is continued until over a finite number of attempts no further minimization of trajectory time is possible. Prior to delineating the algorithm it is to be noted that the set of all collision-free space of the workspace is denoted as C_{free} and the current trajectory of the robot as $\tau_c(s)$. A point of discretization on a trajectory discretized into N parts is denoted as $p(s_i), i \in \{1, 2, \ldots, N\}$. The corresponding configuration of the robot at those points is denoted by $q(s_i)$. The algorithm is given as Algorithm 1.

Algorithm 1 Globally reducing trajectory time

```
$N_{try} \leftarrow 0$
while $N_{try} < N_{attempts}$ do
  Discretize current trajectory $\tau_c(s)$ into $N_p$ parts
    where $N_p$ is selected based on minimum
    discretization distance between two points.
  Set $flag \leftarrow 0$
  for $i = 1$ to $N_p$ do
    Compute minimum velocity at si due to
      shadowing vertices as $v_{rmin}(s_i)$
    if $v_{rmin}(s_i) < v_{rm}$ then
      Find a configuration $q(s_p) \in C_{free}$ and
        $s_p \in \tau c(s_k), k \in \{1, \ldots, N_p\}$ such that $q(s_p)$
        is reachable from $q(s_i)$.
      Find a point $s_r$ on the remaining part of
        the trajectory, $s_r \in \tau_c(s_j); i < j \leq N_p$ such
        that $q(s_r)$ is reachable from $q(s_p)$.
      Form a new trajectory through $s_i, s_p, s_q$
        and denote it as $\tau_n(s)$
      if $T(\tau_n) < T(\tau_c)$ then
        discretize $\tau_n$ into $N_q$ points.
        $\tau_c \leftarrow \tau_n$
        $N_p \leftarrow N_q$
        Set $flag \leftarrow 1$
      end if
    end if
  end for
  if $flag = 0$ then
    $N_{try} \leftarrow N_{try} + 1$
  end if
end while
```

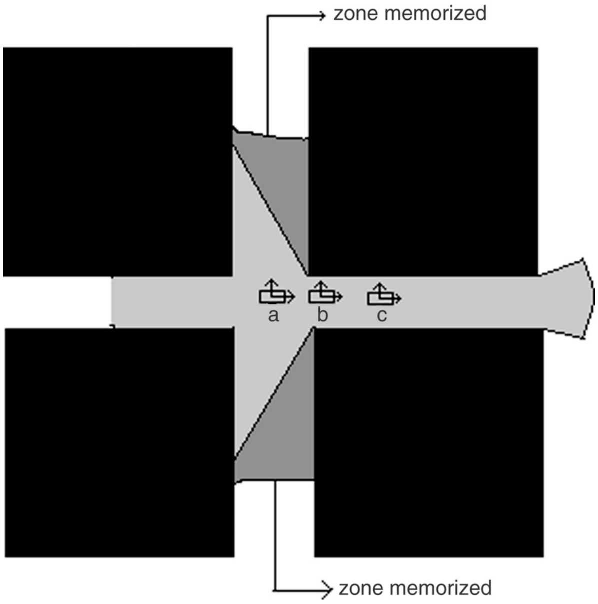


Fig. 8. Memorization of previous scenes.

Step 8 of the algorithm is carried out by searching for a collision-free configuration which would displace the path away from the shadowing vertex responsible for the lowest velocity at s_i. Step 11 adapts the displaced path as the new current path if its trajectory time is less than that of the current path. N_{attempts} is the number of unsuccessful attempts at minimizing the trajectory time before the algorithm halts.

3.5. Remembering sensor information

The computation of the velocity profile at a given point on the robot's trajectory incorporates the robot's field of vision at that point. This field can change appreciably between two successive instances of computation. For example, in Fig. 8 the robot at position a has full field of vision of the corridor that is transverse to the robot's trajectory. However, at position b the robot is blind to the zone shown in a darker shade of grey. Hence it needs to slow down as it moves further down to c since it envisages the possibility of a moving object approaching it from the corners of the stationary objects. These corners are the starting areas of the robot's blindzone at b.

However, if the robot could remember the earlier scene it could use this when computing its velocity profile during execution of the planned path. In such a case, if the robot did not see any moving objects in close proximity at a it can make use of this information at b to have a velocity profile from b that is greater than the one computed in the absence of such information. Fig. 8 shows (in darker shade) the zone remembered by the robot. The contour of the remembered area represents the blindzone of the robot at b, from where mobile objects can emanate. The area in a lighter shade of grey is the visibility polygon for the robot at b. With the passage of time the frontier of the remembered area shrinks due to the advancement of the imagined mobile objects from the initial frontier. The details of this scheme are given below.

Remembering is fruitful when a non-shadowing vertex begins to cast a shadow, thereby hiding regions which were previously visible. The set of all vertices that are currently visible, shadowing and were at some prior instant visible, non-shadowing is denoted by $Vsns$. For every vertex $ve \in Vsns$ a corresponding vertex is associated and called the blind vertex. The blind vertices are of three categories explained in Fig. 9 where the vertex a, non-shadowing for the robot at p, becomes shadowing when the robot is at q. Correspondingly the vertex c of the triangular obstacle which was visible and shadowing when the robot was at p becomes invisible when the robot moves to q. Simultaneously one of the other endpoints of a, namely b, would also become inevitably invisible at q. Vertices like b fall in the second category. If b was already outside C_{vis} at p the intersection of C_{vis} with the segment ab, namely o, is identified as the third category of blind vertex. The set of all such vertices is denoted by Vbs. These vertices are advanced by a distance $v_{ob}\Delta t$ where Δt is the time taken by the robot between p and q to new virtual locations along the line that connects those vertices to a. At q the velocity is computed due to the closest of the vertices in the set Vbs at their virtual locations instead of a, which is otherwise the vertex for which Eq. (2) is computed. Such a trend continues until the distance between the robot to the closest hypothetical vertex is less than the actual distance of the robot to a.

The remembering part of the algorithm is given in Algorithm 2. The set of all visible shadowing vertices is denoted by Vsh.

Algorithm 2 Remembering effects on velocity

```
for each vertex $v_e \in Vsh$ do
    if $v_e \in Vsns$ then
        for each vertex $v_b \in Vbs$ associated with $v_e$ do
            Advance $v_b$ by $v_{ob}\Delta t$
        end for
        Denote the distance from the robot's current location, $s_c$, to the closest of all advanced vertices, $v_{bc}$ as $d_{cvb}$
        if $d(s_c, v_e) < d_{cvb}$ then
            Compute velocity due to the virtual vertex $v_{bc}$ through equation (2)
        else
            Compute velocity due to the actual vertex $v_e$ through equation (2)
        end if
    end if
end for
```

4. From plan to execution

The velocity profile, $v_\tau(s)$, is a sequence of maximum velocities calculated at discretized locations along the trajectory $\tau(s)$. The locations at which the velocity profile at the execution stage is computed are not the same locations as where the profile was computed during planning, due to odometric and

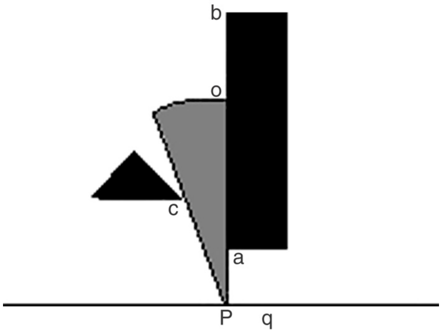


Fig. 9. Three categories of blind vertices.

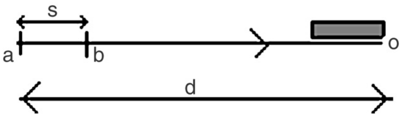


Fig. 10. Effect of an obstacle on the robot's velocity, possibly hiding mobile objects at locations *a* and *b*.

motor constraints. Moreover, if there are changes in the environment it entails modifying the trajectory and hence the velocities. During execution it is computationally expensive to compute the profile for the entire remaining trajectory, hence the profile is computed for the next finite distance, given by $d_{\text{safe}} = d_{\max} + nd_{\text{samp}}$, where $d_{\max} = v_{rm}^2/(2 * a_{-m})$ represents the distance required by the robot to come to a halt while it moves with the maximum permissible velocity afforded by motor constants. And $d_{\text{samp}} = v_{rm}t_{\text{samp}}$ is the maximum possible distance that the robot can move between two successive samples (time instants) of transmitting motion commands, where time between two samples is t_{samp}.

The main issue here is what should be the distance over which the velocity profile needs to be computed during execution such that it is safe. A velocity command is **not** considered safe if it is less than the current velocity and **not** attainable within the next sample. The velocity is constrained by the environment as well as robot's own dynamics and hence their roles are studied below.

Effect of environment

Mobile objects that can emerge from corners in a head-on direction cause the greatest change in velocity over two samples. Fig. 10 shows one such situation, where the rectangular object casts a shadow and is susceptible to hiding mobile objects. Let the current velocity of the robot at a due to the object be v_1. Let the velocity at a distance, s, from a, at b (Fig. 10) due to the object be v_2.

The velocities at a and b are given by

$$v_a = -v_{ob} + \sqrt{v_{ob}^2 + 2a_{-m}d}, \tag{8}$$

$$v_b = -v_{ob} + \sqrt{v_{ob}^2 + 2a_{-m}(d-s)}. \tag{9}$$

Hence

$$v_a^2 - v_b^2 = 2a_{-m}s + 2v_{ob}\left(\sqrt{v_{ob}^2 + 2a_{-m}(d-s)} - \sqrt{v_{ob}^2 + 2a_{-m}d}\right). \tag{10}$$

Evidently the second term on the right-hand side of Eq. (10) is negative, since the second square root term is more positive than the first. Hence $v_a^2 - v_b^2 \le 2a_{-m}s$. Therefore the velocity at b, v_b can be attained from the velocity at a, v_a under maximum deceleration, d_m, irrespective of the maximum velocity of the mobile object or the robot's own motor constraints. This was intuitively expected since the robot's velocity at any location is the maximum possible velocity that guarantees immobility before collision; its velocity at a subsequent location permitted by the environment would be greater than or equal to the velocity at the same location obtained under maximum deceleration from the previous location. In other words, for safeness of velocity going purely by environmental considerations it would suffice to calculate the velocity, for the next sampling distance alone, for without loss of generality, $d = d_{\text{samp}}$.

Effect of robot's dynamics

The robot needs to respect the velocity constraints imposed while nearing the clothoidal arcs and eventually while coming to the target. The robot can reach zero velocity from its maximum velocity over a distance of $d_{\max}$, computed before. Hence $d_{\max} + d_{\text{samp}}$ represents the safe distance over which the velocities need to be computed.

4.1. Online path adaptation for better trajectory time

The third of the problems outlined in Section 2 is tackled here. During navigation the robot in general comes across objects hitherto not a part of the map. The robot reacts to these new objects in line with the basic philosophy of safety as well as time reduced paths. The adaptation proceeds by finding locations over a finite portion of the future trajectory where drops in velocity occur and pushing the trajectory away from those vertices of the objects that caused these drops to areas in free space where higher velocities are possible. A search is made through the newly found locations of higher velocities for a time-reduced path.

Generalized procedure

The generalized procedure for adapting the path in the presence of new objects is delineated through Fig. 11.

1. On the trajectory segment that is currently traversed, AB in Fig. 11, enumerate the vertices of objects that reduce the velocity of the robot.
2. The positions are found on AB where the influence of vertices is likely to be maximal.
3. These positions are pushed by distances $d_p = k(v_l - v_r)$, where v_l and v_r are the velocities at that location on the path due to the most influential vertices on the left and right of

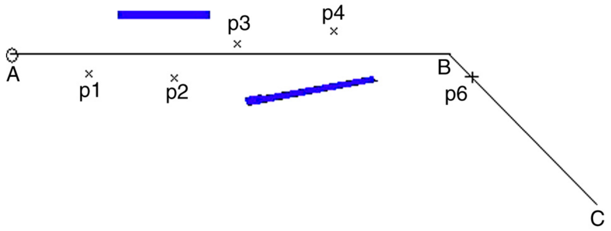


Fig. 11. A trajectory in the presence of new objects. The points marked with crosses represent locations through which a path is searched for reduced time of trajectory.

the path. These new locations are denoted as $p1$, $p2$, $p3$, $p4$ (Fig. 11) and maintained as a list provided the velocity at the new locations is higher than the original ones. $p6$ is the farthest point on the robot's trajectory visible from its current location at A.

4. On this set of locations A, $p1$, $p2$, $p3$, $p4$, $p5$, $p6$ starting from the current location at A, find a trajectory sequence shorter in time than the current sequence of A, B, $p6$ if it exists.
5. The steps 1 to 4 are repeated until the robot reaches the target.

It should be noted that when a collision with an object is detected, a collision-free location is first found that connects the current location with another location on the original trajectory and this new collision-free path is further adapted for a time-reduced path if it exists. Also note that while the velocities are computed over a distance d_{safe}, that part of the remaining trajectory that is visible from the current location is considered for adapting to a better time-length.

5. Planning results and analysis

In this section the results of incorporating the velocity profile computation as a consequence of considering robot and environment dynamics and sensor capacities at the planning stage and the subsequent adaptation of paths to better time of trajectory is analysed. Fig. 12 shows the path computed by a typical holonomic planner [9] and its corresponding velocity profile. The velocity corresponding to the robot's location on the trajectory (shown as a small circle) is marked by a straight line labelled m on the profile. The dark star-shaped polygon centred at the robot depicts the visibility of the robot at that instant and is called the visibility polygon. The figure is a snapshot of the instant when the robot begins to decelerate to a velocity less than half the current velocity as it closes down on the vertex a marked in the figure. Evidently from the visibility polygon the vertex a casts a shadow, and the closer the robot gets to it, the slower the velocity must be.

Fig. 13 is the time-reduced counterpart of Fig. 12. The snapshot is once again at a location close to vertex a. Staying away from a permits nearly maximum velocity. The dip observed in the profile due to vertex a is negligible. Similarly staying away from other vertices such as b allows for a trajectory time of 21.79 s compared to 26.30 s for Fig. 12. Modification of the trajectory for shorter time proceeds along the lines of Section 3.4. For the two examples discussed, the robot's maximum acceleration and deceleration was fixed at 1 m/s^2, maximum velocity at 1 m/s and the sensor range at 7 m. The maximum bound on the object's velocities was 1.5 m/s.

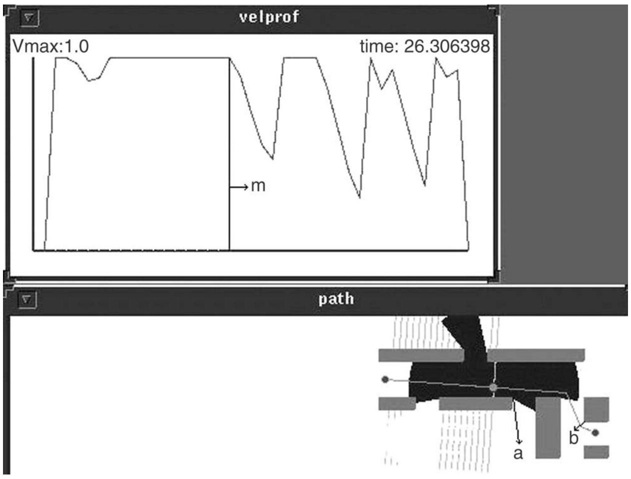


Fig. 12. Path computed by a typical planner and its velocity profile shown on the top. The robot's velocity corresponding to its location on the trajectory is shown by a vertical line on the profile and labelled as m.

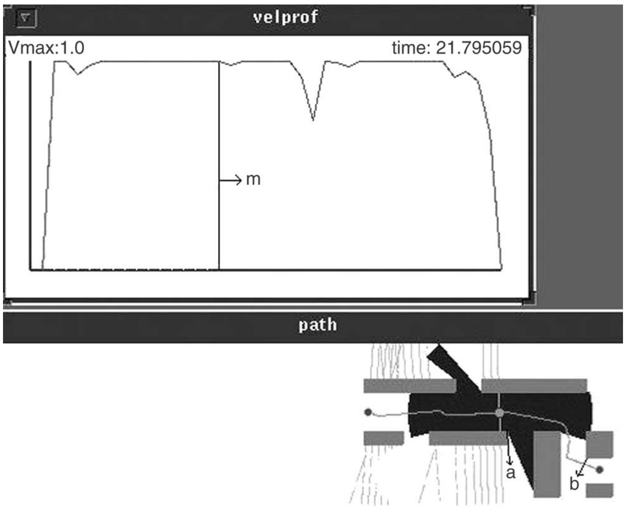


Fig. 13. Path obtained after adaptation to reduced time-length.

Figs. 14 and 15 depict the planned trajectory and velocity profiles before and after reduction of trajectory time for our laboratory environment. The time-reduced trajectory is shorter by more than 8 s as it widens its field of view by moving away from the bends while turning around them.

5.1. *Effect of remembering on trajectory time*

Fig. 16 shows an environment with four corridors named 1, 2, 3 and 4 with planned path obtained by minimizing time. It also portrays the robot's field of vision as it enters corridor 3. The velocity profile for the above path is shown in Fig. 17. The location of the robot corresponding to its location in Fig. 16 is shown through the vertical line. The locations of the robot as it decelerates when its field of view of each of the corridors

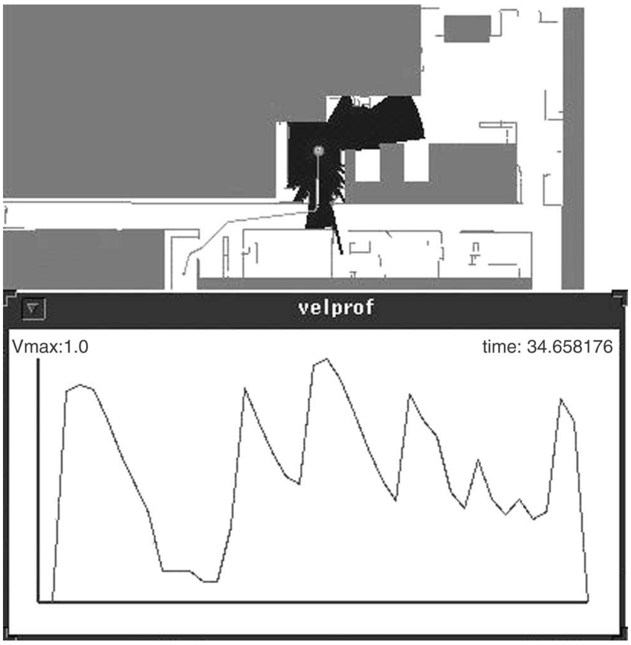


Fig. 14. Planned trajectory before adaptation to a reduced time.

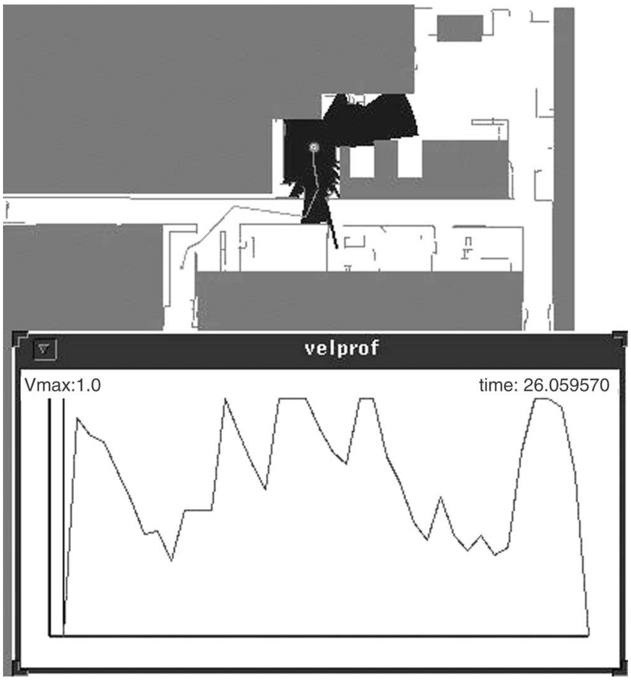


Fig. 15. Time-reduced trajectory at planning stage.

vanishes is also marked with the respective numbers on the profile.

Though the path of Fig. 16 is minimized in time its velocity profile still shows decelerations in the vicinity of the corridors. This is due to the phenomenon discussed in Section 3.5 where the robot becomes blind to many parts of the environment it had seen at the preceding instant. Fig. 18 shows the robot's field of vision at an instant after the one shown in Fig. 16. There is a marked decrease in its field of vision at the latter instant

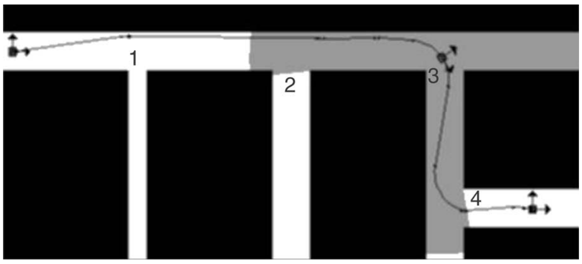


Fig. 16. Robot's field of view as it enters corridor 3.

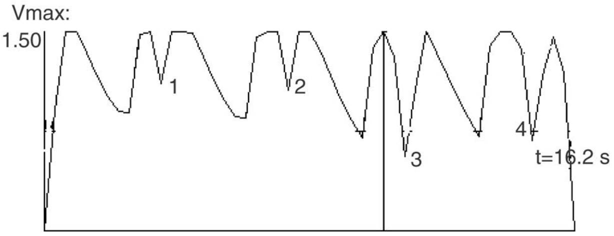


Fig. 17. Velocity profile for the Fig. 16. The corresponding position of the robot is shown as a vertical line. Decelerations near the corridors are also marked with the same numbers.

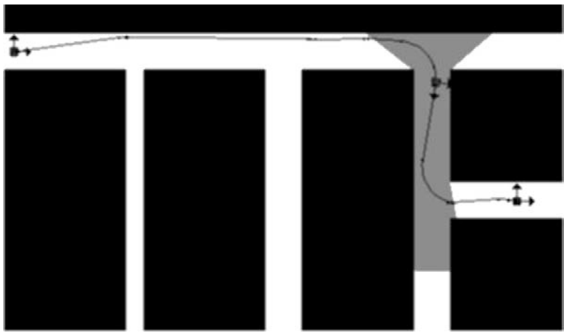


Fig. 18. Robot's field of vision at an instant that immediately follows the instance of Fig. 16.

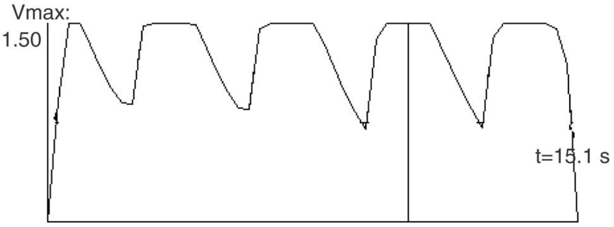


Fig. 19. Velocity profile obtained after incorporation of memory.

that results in the robot reducing its velocity in anticipation of moving objects from the blindzones depicted in the velocity profile.

However, when the robot is able to remember the previous images, the need to decelerate is nullified and the trajectory time is further reduced. Fig. 19 illustrates this where the decelerations shown in the velocity profile of Fig. 17 at locations 1, 2, 3 and 4 are now absent.

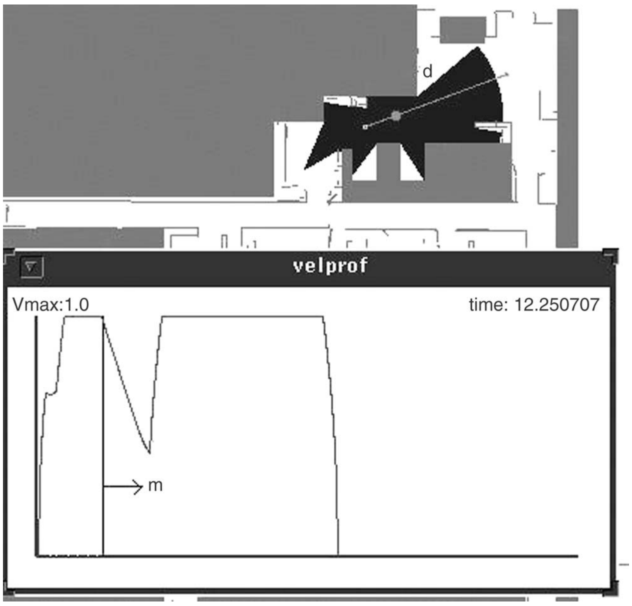


Fig. 20. A simple planned trajectory and its velocity profile.

6. Experimental results

6.1. *Velocity profiles*

In this section the velocity profiles obtained during the planning and execution stages are compared in the absence of any new objects during execution. Fig. 20 shows a simple planned trajectory and the corresponding velocity profile for our laboratory environment. Some of the obstacles are filled in grey and others are shown as segments (in grey). The robot is shown as a small circle and the star-shaped polygon in black represents the field of vision of the robot at that location. The vertical line, marked m, in the velocity profile represents the velocity of the robot corresponding to its position on the trajectory. The profile shows a subsequent drop in velocity, a consequence of the robot getting closer to the region marked d, to which it is blind.

Fig. 21 compares the planned and executed (in simulation) velocity profile. The executed trajectory tallied to a time of 12.28 s in comparison with 12.25 s for the planned profile. These figures illustrate that the executed profiles and execution times are close to the planned profiles and times while there are no changes in the environment.

Figs. 23 and 24 show the execution by the Nomad XR4000 (Fig. 22) of paths computed by a standard planner. Fig. 23 corresponds to the original path computed by the planner and Fig. 24 is its time-reduced counterpart.

The velocity profiles during execution of the two paths are shown in Fig. 25. Some of the bigger drops in the unreduced profile are absent in the reduced profile as the robot avoids turning close to the obstacles that form the bends. The path of Fig. 24 got executed in 12.9 s while the path in Fig. 23 was executed in 13.98 s. The figures are meant as illustrations of the theme that trajectories deformed to shorter time-lengths at planning stage are also executed in shorter time during implementation than their unreduced versions.

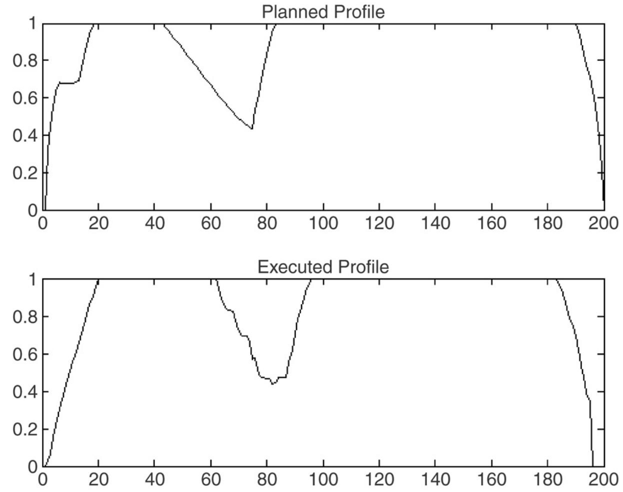


Fig. 21. The planned and executed velocity profile in simulation. The ordinate measures velocity in m/s and the abscissa time in seconds.

Fig. 22. The Nomad XR4000 used in our experiments at LAAS.

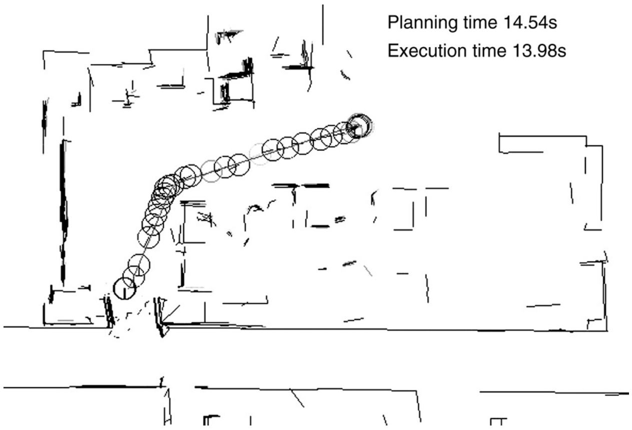


Fig. 23. Execution of the original planned path.

6.2. *Online adaptation of paths for better trajectory time*

This section presents results of the algorithm in the presence of newly added objects that affect the velocities of the robot

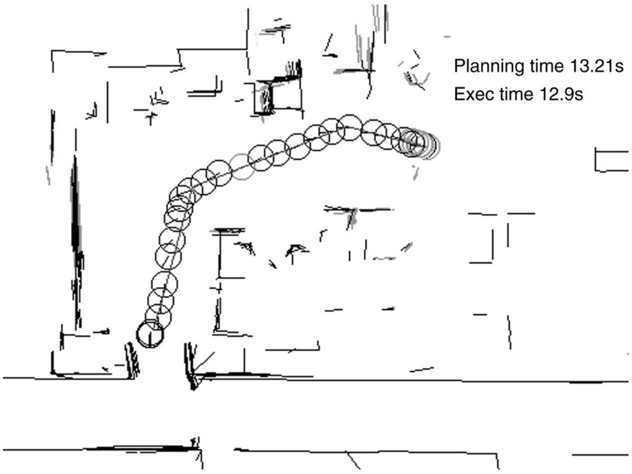


Fig. 24. Execution of the time-reduced path.

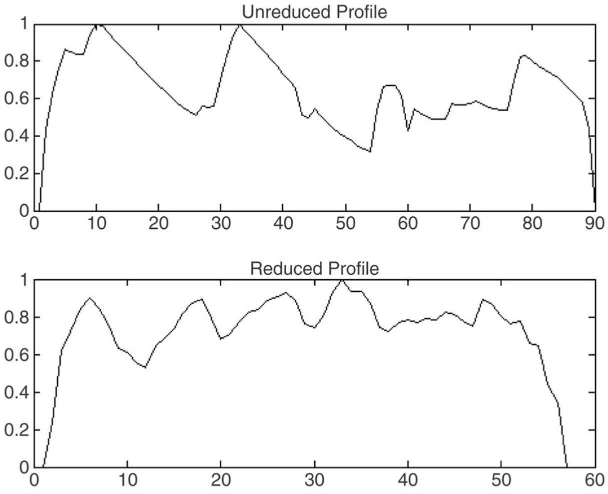


Fig. 25. The top profile corresponds to the path executed in Fig. 23 and the bottom to Fig. 24. The planned and executed velocity profile in simulation. The ordinate measures velocity in m/s and the abscissa time in seconds.

in real time. Fig. 26 shows a path where the robot avoids the two new segments $S1$ and $S2$ intersecting the original planned trajectory but does not adapt its path for better time. The velocity profile for the same is shown in Fig. 27. Fig. 28 is the counterpart of Fig. 29 where the robot adapts its path to a better time-length reactively. The big dips in the velocity profile of Fig. 27 are considerably filtered in Fig. 29 as the robot avoids the obstacles with larger separation. The time-reduced execution tallied to 10.9 s while the unreduced version was executed in 12.5 s. The trajectory time at planning was 7.9 s. The above graphs are those obtained in simulation.

Fig. 30 shows the unreduced executed path by the XR4000 Nomadic robot in our laboratory at LAAS. The obstacles in the original map are shown by black lines, while the segments perceived by the SICK laser are shown in lighter shades of grey. Some of these segments get mapped to the ones in the map and the others are considered new segments. This is done by a segment-based localization algorithm. The segments of concern here are those which form a box-shaped obstacle marked B in

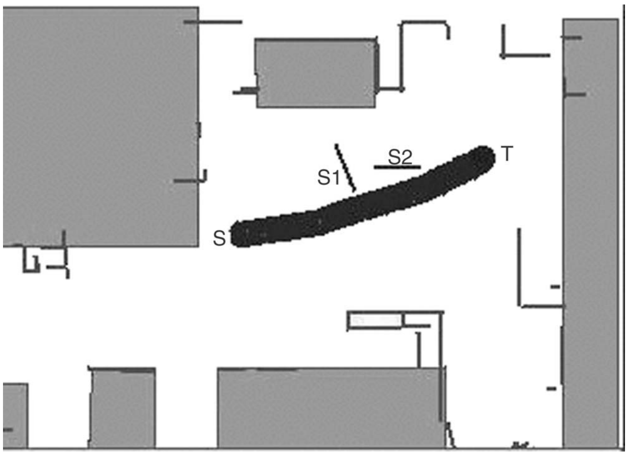


Fig. 26. A simulated execution in the presence of two new segments $S1$ and $S2$ along with the corresponding velocity profile. The path is not adapted to better time-length. Start and goal locations marked as S and T.

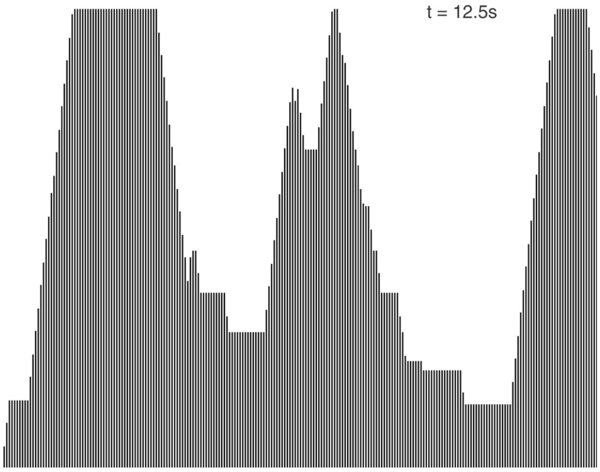


Fig. 27. Velocity profile for the execution of Fig. 26.

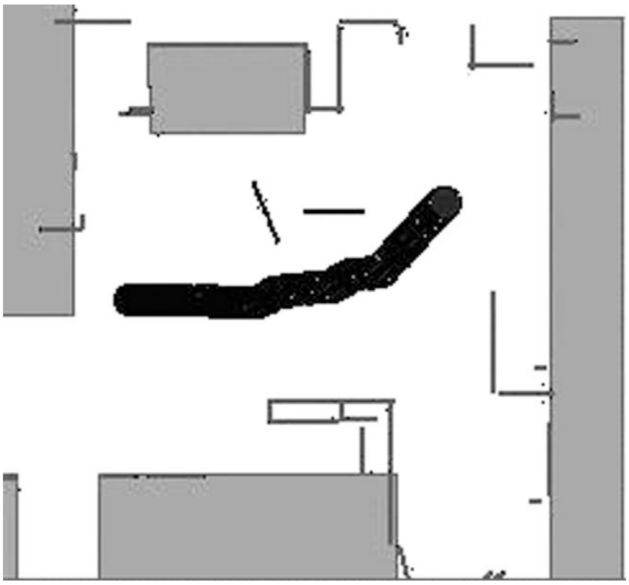


Fig. 28. Path of Fig. 26 adapted to better time-length.

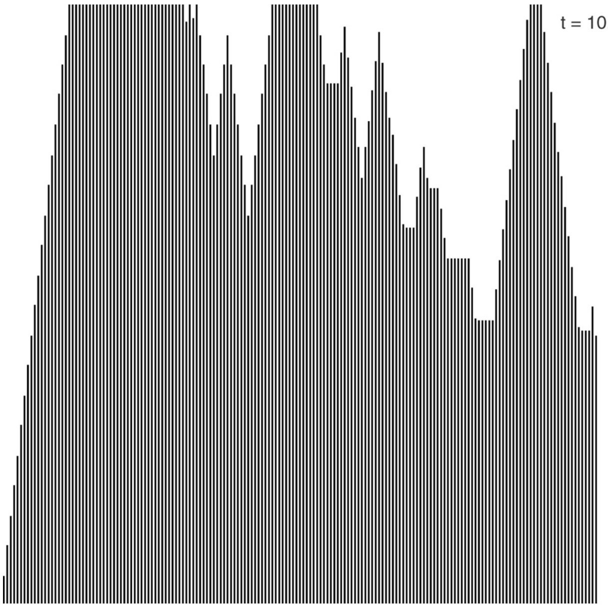


Fig. 29. Velocity profile for the execution of Fig. 28.

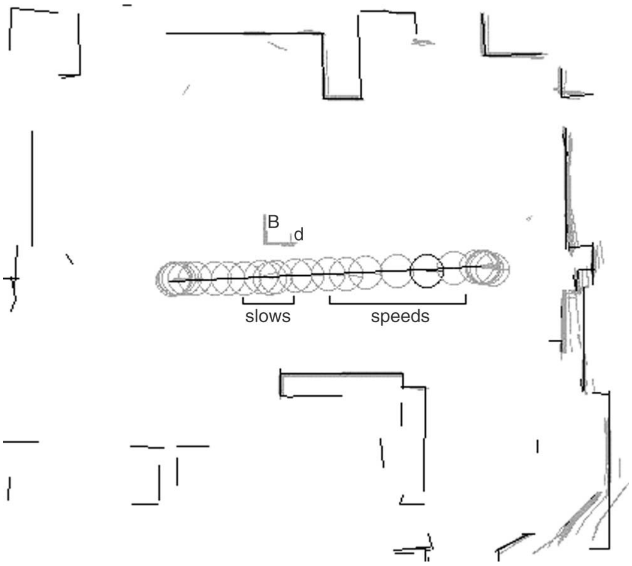


Fig. 30. Unreduced path executed by the Nomad XR4000. The vertex d of the new box-shaped object B forces a slow down near it.

Fig. 30. The vertex d of this obstacle casts a shadow on the robot's sensory field, which forces it to slow down at those locations due to Eq. (2). The execution time for this unreduced path is 10.6 s.

The time-reduced counterpart is shown in Fig. 31 that tallied to 9.6 s. The original planning time was 8.8 s in the absence of the box-shaped object. The corresponding velocity profile is shown in Fig. 32.

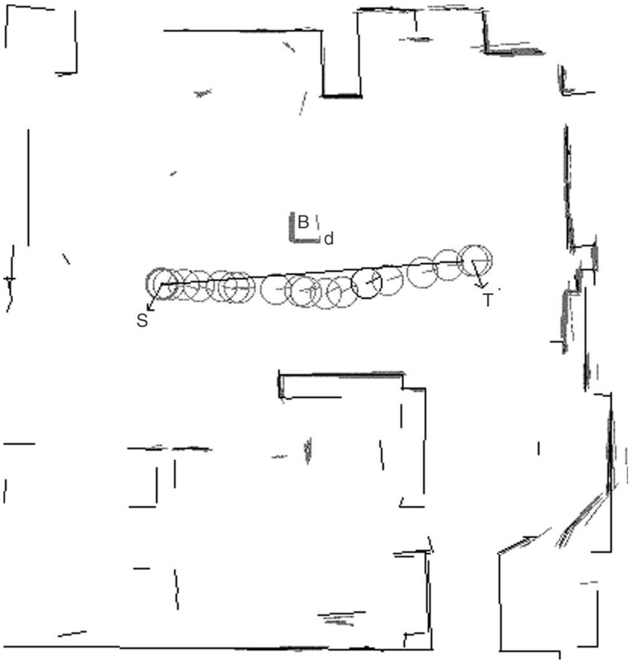


Fig. 31. Time-reduced path executed by the Nomad XR4000. Increasing linear and angular separation from vertex d facilitates a higher speed.

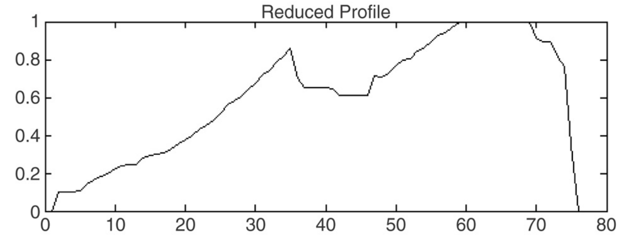


Fig. 32. Velocity profile for the path executed by the Nomad in Fig. 31. The planned and executed velocity profile in simulation. The ordinate measures velocity in m/s and the abscissa time in seconds.

7. Conclusions and scope

A proactive safe planning algorithm and its reactive version that facilitates real-time execution has been presented. The proactive nature of the algorithm stems from the computed velocity profile, $v_\tau(s)$, that guarantees immobility of the robot before collision with any of the possible mobiles that could interfere with its future trajectory from regions blind to its sensor. The proactivity does not however come at the cost of the robot's velocity or trajectory time. The knowledge of $v_\tau(s)$ computed over the trajectory $\tau(s)$ further facilitates reduction of the overall trajectory time $T(\tau)$ by adaptation of the initially planned path. Analysis of the scheme at the planning stage depicts that the robot can have a velocity profile that achieves its maximum possible velocity for a sustained duration without many dips, provided it stays away from doorways and narrow passages along its path. Remembering of previous scenes also enhances the robot's performance through reduced trajectory time and a more uniform velocity profile.

A reactive extension of the scheme that facilitates real-time simulation and implementation is also presented. The scheme maintains the underlying philosophy of computing safe velocities and modification of paths for better trajectory time. Simulation and experimental results at real time corroborate our earlier results obtained at the planning stage (that by keeping away from vertices of objects that could hide mobiles the robot could move at higher velocities and obtain better time-lengths) and thus the efficacy of the overall strategy is vindicated. The minimum distance over which the velocities need to be computed on the remaining trajectory during real time such that the computed velocities are safe is theoretically established. This avoids repetitive computation of velocities over the entire remaining trajectory for every motion command, thereby reducing computational intensity and facilitating real-time implementation. The methodology could be useful in the context of personal robots moving in areas where interference with mobile humans, especially aged ones, is generally expected.

The immediate scope of this work involves incorporating the memory phenomena at the reactive level such that higher speeds are possible. The methodology needs to be validated in the presence of mobile objects that actually impinge on the path from blindzones with a provision for the robot to avoid the objects without halting, and continuing to respect safety considerations as well as minimizing trajectory time.

Acknowledgments

The work described in this paper was conducted within the EU Integrated Project COGNIRON (“The Cognitive Companion”) and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020 and by the French National Program ROBEA.

References

[1] R. Alami, T. Siméon, K. Madhava Krishna, On the influence of sensor capacities and environment dynamics onto collision free motion plans, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, EPFL, Swizerland, 2002.

[2] J.C. Alvarez, A. Skhel, V. Lumelsky, Accounting for mobile robot dynamics in sensorbased motion planning: experimental results, in: IEEE International Conference on Robotics and Automation, Leuven, Belgium, 1998.

[3] B. Bouilly, T. Siméon, R. Alami, A numerical technique for planning motion strategies of a mobile robot in presence of uncertainty, in: IEEE International Conference on Robotics and Automation, Nagoya, Japan, 1995.

[4] D. Cruzel, Planification de mouvements sous contraintes de perception, Master’s Thesis, LAAS-CNRS, 1998.

[5] P. Fiorinin, Z. Schiller, Motion planning in dynamic environments using velocity obstacles, International Journal of Robotics Research 17 (7) (1998) 760–772.

[6] S. Fleury, P. Soueres, J.P. Laumond, Primitives for smoothing mobile robot trajectories, IEEE Transactions on Robotics and Automation 11 (3) (1995) 441–448.

[7] M. Khatib, B. Bouilly, T. Siméon, R. Chatila, Indoor navigation with uncertainty using sensorbased motions, in: IEEE International Conference on Robotics and Automation, Albuquerque, USA, 1997.

[8] K. Madhava Krishna, R. Alami, T. Siméon, Moving safely but not slowly reactively adapting paths for better trajectory times, in: IEEE International Conference on Advanced Robotics, Quimbra, Portugal, 2003.

[9] J.C. Latombe, Robot Motion Planning, Kluwer Academic, 1991.

[10] A. Lazanas, J.C. Latombe, Motion planning with uncertainty: a landmark approach, Artificial Intelligence (1995) 287–315.

[11] J. Minguez, L. Montano, Nearness diagram navigation. a new realtime collision avoidance approach, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, 2000.

[12] N. Roy, S. Thrun, Motion planning through policy search, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, EPFL, Swizerland, 2002, pp. 2419–2425.

[13] Z. Schiller, F. Large, S. Sekhavat, Motion planning in dynamic environments: Obstacles moving along arbitrary trajectories, in: IEEE International Conference on Robotics and Automation, 2001, pp. 3716–3721.

[14] C. Stachniss, W. Burgard, An integrated approach to goal-directed obstacle avoidance under dynamic constraints for dynamic environments, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, EPFL, Swizerland, 2002.

[15] S. Suri, J. O’Rourke, Worst-case optimal algorithms for constructing visibility polygons with holes, in: ACM Symp. on Computational Geometry, 1986.

[16] D. Fox, W. Burgard, S. Thrun, The Dynamic Window Approach to Collision Avoidance, in: IEEE Robotics and Automation Magazine, 1997.

[17] O. Brock, O. Khatib, High speed navigation using the global dynamic window approach, in: IEEE International Conference on Robotics and Automation, 1997.

K. Madhava Krishna obtained his Ph.D. from the Indian Institute of Technology at Kanpur in 2001. He spent the year 2002 at LAAS-CNRS, Toulouse as a post-doctoral fellow with the Robotics and Artificial Intelligence group.

His research interests are in the areas of mobile robotics and multi-robotic systems. He currently directs the Robotics Research Center (iRL) at the International Institute of Information Technology at Hyderabad, India. Prior to this he was a visiting faculty at the CSCE department at the University of Arkansas.

Rachid Alami is a senior research scientist at the Laboratoire d’Automatique et d’Architecture des Systemes, Centre National de la Recherche Scientifique (LAAS-CNRS), Toulouse, France. He joined the Robotics and Artificial Intelligence group at LAAS-CNRS (http://www.laas.fr)in 1980.

His major interests are in the study of control architectures for autonomous robots, task and motion planning, multi-robot cooperation, personal robots and human–robot interaction. He received a “Diplome d’Ingenieur” (1978) and a “Diplome de Docteur-Ingenieur” (1983) in computer science from the Institut National Polytechnique de Toulouse as well as a “Habilitation a Diriger des Recherches” (1996) from University Paul Sabatier, Toulouse, France.

Thierry Simeon graduated from the Institut National des Sciences Appliquees in 1985. He received his Ph.D. degree in Robotics and his Habilitation degree from the University Paul Sabatier at Toulouse, France in 1989 and 1999, respectively. After a postdoctoral year at the University of Pennsylvania, Philadelphia, he joined LAAS-CNRS in 1990 as Charge de Recherche.

His research interests include robotics and algorithmic motion planning. He has published more than 70 papers in international journals and conferences. He has participated in several European projects: the Esprit 3 BRA project PROMotion (Planning RObot Motion), the Eureka project I-ARES (Rover for Planetary Exploration) and more recently the Esprit 4 LTR project Molog (Motion for Logistics) that he coordinated from 2000 to 2002. He is currently involved in the Esprit 5 IST project Movie (Motion in Virtual Environments) and is investigating the application of robot motion planning to computational biology. He is one of the founding partners of Kineo CAM, a spin-off company from LAAS, created in 2001 to develop and market the motion planning technology.

A Methodological Approach relating the Classification of Gesture to Identification of Human Intent in the Context of Human-Robot Interaction*

Chrystopher L. Nehaniv, Kerstin Dautenhahn
Adaptive Systems Research Group
School of Computer Science, University of Hertfordshire
College Lane, Hatfield Herts AL10 9AB, United Kingdom
`C.L.Nehaniv@herts.ac.uk`

Jens Kubacki, Martin Haegele, Christopher Parlitz
Fraunhofer Institute for Manufacturing
Engineering & Automation (IPA)
Nobelstraße 12
70569 Stuttgart, Germany

Rachid Alami
LAAS/CNRS, Robotics & Artificial Intelligence Group
7, Avenue du Colonel Roche
31077 Toulouse Cedex 4, France

***Abstract*— In order to infer intent from gesture, a broad classification of types of gestures into five main classes is introduced. The classification is intended as a generally applicable basis for incorporating the understanding of gesture into human-robot interaction (HRI). Examples from human-robot interaction show the need to take into account not only the kinematics of gesture, but also the interactional context. Requirements for the operational classification of gesture by a robot interacting with humans are suggested and initial steps in its deployment are discussed.**

***Index Terms*— interaction context, classification of gestures, human-robot activity and interaction.**

I. INTRODUCTION: THE NEED FOR CLASSIFYING GESTURE

The word *gesture* is used for many different phenomena involving human movement, especially of the hands and arms. Only some of these are interactive or communicative. The pragmatics of gesture and meaningful interaction are quite complex (cf. [9], [11], [12]), and an international journal [6] now exists entirely devoted to the study of gesture. Applications of service or 'companion' robots that interact with humans, including naive ones, will increasingly require human-robot interaction (HRI) in which the robot can recognize *what* humans are doing and to a limited extent *why* they are doing it, so that the robot may act appropriately, e.g. either by assisting, or staying out of the way. Due to the situated embodied nature of such interactions and the non-human nature of robots, it is not possible to directly carry over methods from human-computer interaction (HCI) or rely entirely on insights from the psychology of human-human interaction. Insights from proxemics and kinesics, which study spatial and temporal aspects of human-human interaction [7], [4], [9] and some insights of HCI, e.g. recognizing the diversity of users and providing feedback acknowledgment with suitable response timing (e.g. [16]), may also prove to be extremely valuable to HRI. Notwithstanding, the nascent field of HRI must develop its own methods particular to the challenges of embodied interaction between humans and robots. New design, validation, evaluation methods and principles particular to HRI must be developed to meet new challenges such as *legibility*, making the robot's actions and behaviour understandable and predictable to a human, and '*robotiquette*', respecting human activities and situations (e.g. not interrupting a conversation between humans or disturbing a human who is concentrating or working intensely — without sufficient cause), as well as respecting social spaces, and maintaining appropriate proximity and levels of attention in interaction. Part of meeting these challenges necessarily involves some understanding of human activity at an appropriate level. This requires the capabilities of recognizing human gesture and movement, and inferring intent. The term "*intent*" is used in this paper in a *limited* way that refers to *particular motivation(s) of a human being that result in a gestural motion directly or indirectly relevant for human-robot interaction*.

In inferring the intent from a human's gesture it is helpful to have a classification of which type of gesture is being observed. Without a sufficiently broad classification, understanding of gesture will be too narrow to characterize what is happening and appropriate responses will not be possible in many cases.

Knowing how to recognize and classify gesture may also serve to inform the design of robot behaviour, including gestures made by the robot to achieve legibility and convey aspects of the robot's state and plans to humans. This in turn will contribute to robot interaction with humans that is legible, natural, safe, and comfortable for the humans interacting with the robot. To begin to approach the complexity of gesture in the context of situated human-

---

*The work described in this paper was conducted within the EU Integrated Project COGNIRON ("The Cognitive Robot Companion") and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020. This paper extends and supercedes C. L. Nehaniv, "Classifying Gesture and Inferring Intent" *Proc. AISB'05 Symposium on Robot Companions: Hard Problems and Open Challenges in Robot-Human Interaction*, The Society for the Study of Artificial Intelligence and Simulation of Behaviour, 2005.

robot interaction, the rough classes of gesture described below are developed in order to provide a broad level of description and the first steps toward a pragmatic, operational definition that could be used by an autonomous system such as a robot to help it (1) to infer the intent of human interaction partners, and (2), as an eventual goal, to help the robot use gestures itself (if possible) to increase the legibility of its behaviour.

II. SOME RELATED WORK ON RECOGNIZING GESTURE AND INTENT

The questions of how gestures are acquired and come to be recognized as meaningful by particular individuals in the course of their development (ontogeny of gesture and its recognition), and conventionalized, elaborated, or lost within particular cultures (evolution of gesture) are large and deep issues, but will not be addressed within the scope in this paper. Psychological/linguistic studies of human gesture use and understanding, related classifications relevant for interaction, language evolution, and language acquisition, e.g. by hearing or deaf children, have all been undertaken (cf. [17]). Such understanding of the development of gesture and its functions may help shed light on gesture in human-robot interaction.

While this paper does not attempt a comprehensive survey of the role and recognition of gesture in human-robot interaction, it does suggest inherent limitations of approaches working with a too narrow notion of gesture, excluding entire classes of human gesture that should eventually be accessible to interactive robots able to function well in a human social environment. Much work with data gloves, typically at a low level for hand gesture recognition for virtual reality or of manipulative grasping has been carried out since the 1990's (e.g. [5], [1]). The important role of gesture for intent communication in human-robot interaction is increasingly being acknowledged, although some approaches still focus only on static hand poses rather than dynamic use of more general types of gesture in context; a survey of hand gesture understanding in robotics appears in [13].

Multimodal and voice analysis can also help to infer intent via prosodic patterns, even when ignoring the content of speech. Robotic recognition of a small number of distinct prosodic patterns used by adults that communicate praise, prohibition, attention, and comfort to preverbal infants has been employed as feedback to the robot's 'affective' state and behavioural expression, allowing for the emergence of interesting social interaction with humans [3]. Hidden Markov Models (HMMs) have been used to classifying limited numbers of gestural patterns (such as grasps or letter shapes) and also to generate trajectories by a humanoid robot matching those demonstrated by a human [2]. Multimodal speech and gesture recognition using HMMs has been implemented for giving commands via pointing, one-, and two-handed gestural commands together with voice for intention extraction into a structured symbolic data stream for use in controlling and programming a vacuuming cleaning robot [8]. Many more examples in robotics exist. Nevertheless, aproaches pursued so far in robotics thus tend to use very limited, constrained, and specific task-related gestural repertoires of primitives, and do not attempt to identify general gestural classes. They have tended to focus on a fixed symbolic set of gestures (possibly an extensible one, in which new gestures can be learned), or focus on only a few representatives from one or two of the gestural classes identified here (e.g. grasps (a subclass of manipulative gestures), or on symbolic and pointing gestures).

Fig. 1. **Gestures with similar kinematics but different functions.** Top row: HELLO (left) an interactional gesture (class 4) is similar to STOP (right) a conventional symbol (class 3). Bottom row: PASS OBJECT (left) is similar and TAKE OBJECT are both multifunctional interactional (class 4) and manipulative (class 1) gestures. Activity and situational context – e.g. stage of interaction and current activity (top row), or location of manipulandum, here a bottle (bottom row) – are be used to disambiguate between such kinematically similar gestures.

III. INSUFFICIENCY OF BODY MODEL FITTING ANALYSES: RELATING CONTEXT TO KINEMATICS.

It should be stressed that a single specific instance of a particular the kind of physical gestural motion could, depending on context and interaction history, reflect very different kinds of human intent. It will not always be possible to infer intent based solely on based the mechanical aspects of human movements (such as changes in joint angles) without taking context into account.

Gestures with identical or near identical human kinematics can be different classes. In general, the kinematic picture alone is not enough to determine the class of a gesture or the human's intent. Examples in shown in figure 1 require contextual information in order to be disambiguated. Relating the context and history of interaction to the kinematics is a key point for recognizing human gestures in HRI. Figure 1 using our classification (see below) illustrates this ambiguity.

IV. CLASSIFICATION OF GESTURES

To approach this problem, a classification of gesture for inferring intent and assisting in the understanding of human activity should closely relate gesture with limited categories of intent in situated human activity. The categories of the broad classification presented here thus correspond to and allow the attribution of limited kinds of intent to humans. This classification is developed as an aid for helping robots to achieve limited recognition of situated human gestural motion, so as to be able to respond appropriately if required, while these robots are working in an environment of ambient human activity (such as a home or office), in which, at times, the robots are also assisting or cooperating with the humans. *Applications of this classification will require the mapping of physical aspects of gestural motion in interactional contexts to the five gestural classes (and their subtypes) suggested here.*

The following is a rough, tentative classification. Gestures are classed into five major types with some subtypes.

A. Five Classes (with Subtypes)

1) **'Irrelevant'/Manipulative Gestures.** These include *irrelevant gestures*, *body / manipulator motion*, *side-effects of motor behaviour*, and *actions on objects*. Broadly characterized, manipulation by a human is here understood as *doing something to influence the non-animate environment or the human's relationship to it (such as position).* Gestural motions in this class are manipulative actions (in this sense) and their side effects on body movement. These 'gestures' are neither communicative nor socially interactive, but instances and effects of human motion. They may be salient, but are not movements that are primarily employed to communicate or engage a partner in interaction. Cases include, e.g. motion of the arms and hands when walking; tapping of the fingers; playing with a paper clip; brushing hair away from the face with the hand; scratching; grasping a cup in order to drink its contents. (Note: it may be very important to distinguish among the subtypes listed above for robot understanding of human behaviour.)
2) **Side Effect of Expressive Behaviour.** In communicating with others, motion of hands, arms and face (changes in their states) occur as part of the overall communicative behaviour, but without any specific interactive, communicative, symbolic, or referential roles (cf. classes 3-5)

   Example: persons talk excitedly raising and moving their hands in correlation with changes in voice prosody, rhythm, or emphasis of speech.
3) **Symbolic Gestures.** Gestural motion in symbol gesture is a *conventionalized signal in a communicative interaction.* It is generally a member of a limited, circumscribed set of gestural motions that have specific, prescribed interpretations. A symbolic gesture is used to trigger certain actions by a targeted perceiver, or to refer to something or substitute as for another signal according to a code or convention. Single symbolic gestures are analogous to discrete actions on an interface, such as clicking a button.

   Examples: waving down a taxi for it to stop; use of a conventional hand signals (a command to halt indicated open flat hand; a military salute); nodding 'yes'; waving a greeting 'hello' or 'goodbye'.

   Note that the *degree of arbitrariness* in such gestures may vary: The form of the gesture may be an arbitrary conventional sign (such as a holding up two fingers with palm forwards to mean 'peace', or the use of semaphores for alphabetic letters). On the other hand, a symbolic gesture may resemble to a lesser or greater extent iconically or, in ritualized form, a referent or activity.

   Further examples: holding up two fingers to indicate 'two'; opening both (empty) hands by turning palms down to indicate a lack of something. Nearly all symbolic gestures are used to convey *content* in communicative interactions.
4) **Interactional Gestures.** These are gesture used to *regulate interaction with a partner*, i.e. used to initiate, maintain, invite, synchronize, organize or terminate a particular interactive, cooperative behaviour: raising a empty hand toward the partner to invite the partner to give an object; raising the hand containing an object toward the partner inviting them to take it; nodding the head indicating that one is listening. The emphasis of this category is neither reference nor communication but on gestures as mediators for cooperative action.[1] Interactional gestures thus concern regulating the form of interactions, including the possible regulation of communicative interactions but do not generally convey any of the content in communication. Interactional gestures are similar to class 1 manipulative gestures in the sense that they influence the environment, but in contrast to class 1, they influence the "animated environment" – *doing something to influence human agents (or other agents) in the environment*, but not by conveying symbolic or referential content.[2]
5) **Referential/Pointing Gestures.** These are *used to refer to or to indicate objects (or loci) of interest* – either physically present objects, persons, directions or locations the environment – by pointing (*deixis* – showing), or indication of locations in space being used as proxies to represent absent referents in discourse. Deictic gesture can involve a hand, finger,

---

[1] Note that we are using the word "cooperative" in a sense that treats regulating communication or interaction as an instance of cooperation.

[2] Some more subtle examples include putting one's hand on another person's arm to comfort them. Such actions, and others involving physical contact, may be quite complex to interpret as understanding them may require understanding and modeling the intent of one person to influence that state of mind of another. At this point, we class simply them with interactional gestures recognizing that future analysis may reveal deep issues of human-human interaction and levels of complexity beyond the rudimentary types of human intent considered here. A special case worthy of note is human contact with the robot, unless this is directly a manipulation of the robot's state via an interface - e.g. via button presses — which would fall into class 3 (symbolic gesture), non-accidental human contact with the robot is likely to be indicative of an intent to initiate or regulate interaction with the robot (class 4). Physical contact between humans might also involve expression of affection (kissing), or aggression (slapping, hitting) – which generally indicate types of human-human interaction it would be better for a robot to steer clear of!

other directed motion, and/or eye gaze. Checking the eye gaze target of an interaction partner is commonly used to regulate reference and interaction.[3]

Table I summarizes the five classes.

Data on the interaction history and context may help in determining the class of a gesture. If the class is known, then the set of possible gestures can remain large, or be narrowed significantly. Symbolic gestures (class 3) correspond to discrete symbols in a finite set, of which there may be only be a small number according to context or size of the given repertoire of the given symbolic gestural code. Interactional gestures (class 4) are likely to comprise a small, constrained class. Class 1 gestures are either "irrelevant", or to be understood by seeking the intent of the associated motor action or object manipulation (e.g. grasping or throwing an object, arms moving as a side effect of walking). Class 5 (referential and pointing gestures) comprise a very limited class, although pointing can also at times carry affective force (e.g. hostility).

Knowledge of specific conventional codes and signs can help the identification of particular signs within class 3, and also in determining that the gesture in fact belongs to class 3, i.e. is a symbolic communicative signal. Machine learning methods such as Hidden Markov Models may be used successfully to learn and classify gestures for a limited finite set of fixed gestures (e.g. [18]). It seems likely that HMM methods would be most successful with class 3 (symbolic gestures) or within narrow domains within other classes (manipulative grasps with class 1), but how successful they would be at differentiating between classes or for whole classes remains uninvestigated at present.

V. IMPORTANT ISSUES

A. *Target and Recipient of a Gesture*

If a gesture is used interactively or communicatively (classes 2-5), it is important to recognize whether the gesture is directed toward the current interaction partner (if any) — which may the robot, another person (or animal) present in the context, or possibly neither (*target*). If pointing, what is the person pointing to? Who is the pointing designed to be seen by? (*recipient*). If speaking, to whom is the person speaking? If the gesture is targeted at or involves a contact with an object, this suggests it may belong to class 1 (or possibly 5, even without contact). A gesture of bringing an object conspicuously and not overly quickly toward an interaction partner is manipulative (in the sense explained in the discussion of class 1, since an object is being manipulated), but it may well at the same time also be a solicitation for the partner to take the object (class 4). Similarly if the partner has an object, an open hand conspicuously directed toward the partner or object may be a solicitation for the partner to give the object (class 4).

B. *Multipurpose Gestures*

It is possible for a single instance of a particular gesture to have aspects of more than one class or to lie intermediate between classes. As mentioned above, handing over an object is both class 1 and 4. And, for example, holding up a yellow card in football has aspects of classes 1 and 3, object manipulation and conventional symbolic signal. Many ritualized symbolic gestures (class 3) also can be used to initiate or regulate interaction (class 4), e.g. the 'come here' gesture: with palm away from the recipient, moving the fingers together part way toward the palm; waving forearm and open hand with palm facing recipient to get attention. More complex combinations are possible, e.g. a gesture of grasping designed by the human to be seen by a recipient interaction partner and directed toward a heavy or awkwardly-sharped target object as a solicitation of the partner to cooperatively carry the object with the gesturer (classes 1, 4, 5).

C. *Ritualization: Movement into Classes 3 and 4*

Gestures that originate in class 1 as manipulations of the non-animate environment and the person's relationship to it may become *ritualized* to invite interactions of certain types, e.g., cupping the hand next to the ear can indicate that person doing it cannot hear, so that the interaction partner should speak up. Originally cupping the hand near the ear served to improve a person's ability to hear sounds in the environment from a particular direction (class 1), but it may be intended to be seen by a conversational partner who then speaks up (class 4). The hand cupped at the ear can even be used as a conventionalized symbol meaning 'speak up' (classs 3). Other examples of ritualization toward regulation of interaction and also symbolic gesture include mimicking with two hands the motions of writing on a pad as a signal to a waiter to ask for the bill; miming a zipping action across the mouth to indicate that someone should be 'shut up'; or placing a raised index finger over lips which have been pre-formed as if to pronounce */sh/*.

D. *Cultural and Individual Differences*

Different cultures may differ in their use of the various types of gesture. Some symbolic gestures such as finger signs (e.g. the "OK" gesture with thumb and index finger forming a circle) can have radically different interpretations in other cultures, or no set interpretation depending on the culture of the recipient (e.g. crossing fingers as a sign of wishing for luck, or the Chinese finger signs for some numbers such as 6, 7, 8). Tilting the head back (Greece) or nodding the head (Bulgarian) are used symbolically for 'no', but would certainly not be interpreted that way in many other cultures. Cultures also differ in their types and scope of movement in (class 2) expressive gestures: Consider, for example, the differences of rhythm, prosody, hand motions, eye contact, and facial expressions accompanying speech between British, Italian, Japanese, and French speakers.

Within cultures, differences between different individuals' uses of gestures can be regional, restricted to particular

---

[3] Eye gaze following develops and supports joint attention already in preverbal infants. Language, including deictic vocabulary (e.g. demonstratives such as the words "these" and "that"), and other interactional skills, typically develop on this scaffolding (see [10]).

Classification of gestural classes and associated (limited) categories of human intent

Class	Name	Defining Characteristics and Associated Intent
1	'Irrelevant' and Manipulative Gestures	Influence on non-animate environment or human's relationship to it; manipulation of objects, side effects of motor behavior, body motion
2	Side Effect of Expressive Behaviour	Expressive marking, (no specific direct interactive, symbolic, referential role) associated to communication or affective states of human
3	Symbolic Gestures	Conventionalized signal in communicative interaction; communicative of semantic content (language-like)
4	Interactional Gestures	Regulation of interaction with a partner; influence on human (or other animated) agents in environment but generally with lack of any symbolic/referential content used to initiate, maintain, regulate, synchronize, organize or or terminate various types of interaction
5	Referential/Pointing Gestures	Deixis; indicating objects, agents or (possibly proxy) loci of discourse topics, topics of interest; pointing of all kinds with all kinds of effectors (incl. eyes): referential, topicalizing, attention-directing

TABLE I

Five Classes of Gesture. See text for explanation, details and examples. Note that some occurrences of the same physical gesture can be used in different classes depending on context and interactional history; moreover, some gestures are used in a manner that in the same instance belongs to several classes (see text for examples).

social groups within the culture, and vary in particularities (such as speed, repertoire, intensity of movement, etc.) between individuals according to preference or ontogeny. Elderly and young may employ gestures in different ways.

VI. Inferring the Intent of Gesture

Being able to identify details of gestural kinematics and even to classify into one of the above classes gives us only starting points for inferring the intent of the person making the gesture due to frequent ambiguity. Resolving this points to the important roles of context and interactional history. Thus, it is necessary to develop operational methods for recognizing the class of gesture in a particular context.[4] If the interactional context of recent activity in which a gesture occurs is known, this can suggest possibilities for which classes (and subtypes) of gesture might be involved. Information on the state of human (e.g. working, thirsty, talking, ...) often can limit the possibilities. Data on the following could help the robot classify the gesture and infer the intent of the human:

(a) the activity of the gesturer is known,
(b) previous and current interaction patterns are remem-

---

[4] Knowledge of the immediate context in some cases needs to be augmented by taking into account of the broader *temporal horizon* of interactional history (cf. [14]).

bered to predict the likely current and next behaviour of the particular person,

(c) objects, humans and other animated agents in the environment are identified and tracked.

(d) the scenario and situational context are known (e.g. knowing whether a gesture occurs at a tea party or during a card game).

A programme to apply the above classification can be developed as follows: (1) Identify the many, particular gestural motions that fit within each of the five classes. Some gestural motions will appear in more than one class. For example, the same mechanical motion of putting a hand and arm forward with the forearm horizontal and the hand open could indicate preparation to manipulate an object in front of the human (class 1), to show which object is being referred to (class 5), or to greet someone who is approaching, or to ask for an object to be handed over (both class 4). (2) Gestural motions identified as belonging to several classes need to be studied to determine in which contexts they occur: determining in which class(es) particular a instance of the gesture is being used may require consideration of objects and persons in the vicinity, the situational context, and the history of interaction. (3) Systematic characterizations of a physical gestural motion together with interactional contexts in which they are occur could then be used to determine the likely class. (4) Deploy on-board characterizations of the relationships between classes and kinematic gestural motions for a range of typical interactional contexts to infer intent and guide robot behaviour. (5) Updating the Interaction History: Attribution of intent related to gesture can then feedback into understanding of the situational context, including motivational state of the human performing the gesture, and becomes part of the updated interaction history, which can then help in inferring intent from ensuing gestures and activity.

VII. HEURISTIC-BASED FAST RECOGNITION AND DISAMBIGUATION OF GESTURES WITH A TIME-OF-FLIGHT DEPTH SENSOR

There are mainly three methods to recognize and detect gestures: model-based approaches fit a kinematics model into the scene observed by sensors, recognition based on classifiers use learning algorithms to label gestures, and heuristic-based methods which directly search for hints related to a gesture. Depending on the context of the overall robotic control system, all of these may be of use. The model-based approach is followed by researchers in the Cogniron project, and, as shown above, this must be augmented by contextual and situational knowledge. The goal is to develop algorithms that geometrically fit a model maintained by the robot into the current scene observed by stereo vision systems and a time-of-flight depth sensor proposed in [15]. Apart from this exhaustive approach there is also work related to a computationally much cheaper heuristic-based method only using data delivered by the depth sensor. The motivation for this is two-fold. Firstly, a 'quick' check of the existence of humans in the close vicinity of the robot and a first basic evaluation of possibly important gestures can be used directly for communication. Secondly, outputs of a fast algorithm related for instance to body, head or arm positions can serve to trigger more detailed investigation by e.g. model-based algorithms. Additionally, the data can be used to initialize model fitting.

The heuristic approach first divides the depth scene observed by the sensor into consecutive depth intervals each having a fixed distance and size. The two intervals containing the most measurements are used for binary segmentation of the humans profile. Within the profile the algorithm searches for a human's center point by summing all pixels belonging to the profile and averaging their coordinates. From this point the algorithm searches upwards and determines a bounding box for the head including the neck by incorporating estimates of the shoulder end points. They can be found as being the left and right extremes of the profile at a height roughly at the bottom of the head. Interestingly, the height of the bounding box around the head plus the width of the shoulder can give an estimate of the length of the upper arm as described in medical statistics. Based on this information four cases can be distinguished:

- Outstretched arm away from the body
- Outstretched arm up or down
- Bent arm next to body
- Bent arm in front of body

For each case further heuristic algorithms are used to determine the hand position and orientation. This can be used to recognize and discriminate between basic gestures. See Figure 2 for a visualization of how the program finds a WAVE gesture and SHAKE HANDS gesture; both are interactional gestures (class 4). By using additional information such as orientation and distance of the human towards the robot and internal state of the robot, tentative disambiguations between similar gestures have been made.

VIII. CONCLUSIONS, NEXT STEPS AND THE FUTURE

In order to infer the intent of a human interaction partner, it may be useful to employ a classification of gesture according to some major types – five in the tentative classification proposed here – whose intent may be (1) absent / directed to objects or environment, (2) incidentally expressive, (3) symbolic, (4) interactional, or (5) deictic. A summary of the classes is given by Table I.

In order to deploy the inference of intent on robots interacting with humans it is necessary to operationalize the distinctions between these (sometimes overlapping) classes. This may require the use of knowledge of human activity, recognition of objects and persons in the environment, and previous interactions with particular humans, as well as knowledge of conventional human gestural referencing and expression, in addition to specialized signaling codes or symbolic systems.

Work in Cognrion now focuses on the organization of the robot decisional abilities and more particularly on the management of human interaction. There is explicit

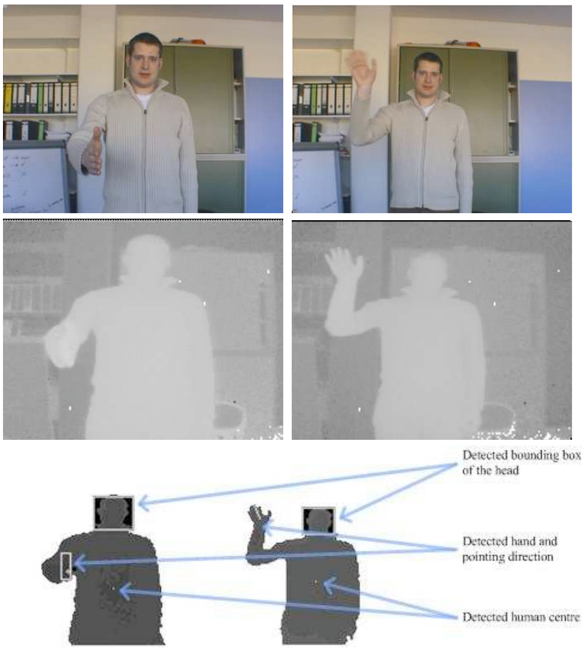


Fig. 2. **Heuristic-based fast recognition and disambiguation of interactional gestures.** Detecting bounding box of head, hand and its orientation, and center of human (top row: color image of gesturing human in robot's vicinity. second row: depth image derived from camera images. Left column: SHAKE HANDS gesture, right column: WAVE gesture.

management of the interactions between the robot companion and its human partners. This requires essentially task-oriented processes that each consist of establishing a common goal, achieving it and verifying commitment of all agents involved during the task performance. Indeed, perception of the human partner is one essential source of information all along the human-robot interaction process from the detection of human presence to the monitoring of human activity and the continuous estimation of its commitment level to a joint goal. This viewpoint is compatible with and served by the classification of gestures proposed here. It also helps us to operationalize use of the classification. Indeed, gestures of type 3 and many of type 1 may be considered as task-oriented and the inference of their intent can be done relative to the task at hand. Gestures of type 4 include generic interactional gestures that may serve to manage the session itself: inviting the robot to start an interaction, suspending or stopping an interaction session, etc. Many gestures of type 4 are consequently task independent.

The classification presented here suggests some requirements for the design and implementation of systems inferring intent from gesture based on this classification. These requirements might be realized in a variety of different ways using, e.g. continuous low-key tracking or more detailed analysis, event-based and/or scenario-based recognition, and prediction of human activity based on models of human activity flows (with or without recognition of particular humans and their previous interactions), depending the particular needs of the given human-robot interaction design and the constraints and specificity of its intended operational context. Design of a robot restricted to helping always the same user in the kitchen environment would be quite different from one that should be a more general purpose servant or companion in a home environment containing several adults, children and pets, but the classification presented here is applicable in informing the design of gesture recognition for inferring intent in either type of system, and for designing other HRI systems.

Finally, effective human-robot interaction will require generation of gestures and feedback signals by the robot. The classification given here can suggest categories of robotic gestures that could be implemented to improve the *legibility* to humans of the robot's behaviour, so that they will be better able to understand and predict the robot's activity when interacting with it.

REFERENCES

[1] BERNADIN, K., OGAWARA, K., IKEUCHI, K., AND DILLMANN, R. A sensor fusion approach for recognizing continuous human grasping sequences using hidden markov models. *IEEE Trans. on Robotics 21*, 1 (2005), 47–57.

[2] BILLARD, A., EPARS, Y., CALINON, S., CHENG, G., AND SCHAAL, S. Discovering optimal imitation strategies. *Robotics & Autonomous Systems,* Special Issue: Robot Learning from Demonstration *47*, 2-3 (2004), 69–77.

[3] BREAZEAL, C., AND ARYANANDA, L. Recognition of affective communicative intent in robot-directed speech. *Autonomous Robots 12*, 1 (2002), 83–104.

[4] CONDON, W. S., AND OGSTON, W. D. A segmentation of behavior. *Journal of Psychiatric Research 5* (1967), 221–235.

[5] DUBOIS, E., NEDEL, L. P., FREITAS, C. M. D. S., AND JACON, L. Beyond user experimentation: notational-basd systematic evaluation of interaction technieuqs in virtual reality environments. *Virtual Reality 8*, 2 (2005), 118–128.

[6] *Gesture*. ISSN: 1568-1475. John Benjamins Publishing Co., The Netherlands `http://www.benjamins.com/cgi-bin/t_seriesview.cgi?series=GEST`.

[7] HALL, E. T. *The Dance of Life: The Other Dimension of Time*. Anchor Books, 1983.

[8] IBA, S., PAREDIS, C. J. J., AND KHOSLA, P. K. Interactive multimodal robot programming. In *Proceedings of the 2002 IEEE International Conference on Robotics and Automation, Washington D.C., May 11-15, 2002* (2002).

[9] KENDON, A. Movement coordination in social interaction: Some examples described. *Acta Psychologica 32* (1970), 100–125.

[10] KITA, S., Ed. *Pointing: Where Language, Culture and Cognition Meet*. Lawrence Erlbaum Associates, Inc, 2003.

[11] MEY, J. *Pragmatics: An Introduction*. Blackwell Publishers, 2001.

[12] MILLIKAN, R. G. *The Varieties of Meaning: The 2002 Jean Nicod Lectures*. MIT Press/Bradford Books, 2004.

[13] MINERS, W. B. *Hand Gesture for Interactive Service Robots*. MSc Thesis, The University of Guelph, Faculty of Graduate Studies, August 2002.

[14] NEHANIV, C. L., POLANI, D., DAUTENHAHN, K., TE BOEKHORST, R., AND CAÑAMERO, L. Meaningful information, sensor evolution, and the temporal horizon of embodied organisms. In *Artificial Life VIII* (2002), MIT Press, pp. 345–349.

[15] OGGIER, T., AND LANG, G. K. User manual: Swissrangertm 2, December 2003, Rev. March 2004.

[16] SHNEIDERMAN, B. *Designing the User Interface: Strategies of Effective Human-Computer Interaction*, 3rd ed. Addison-Wesley, 1998.

[17] VOLTERRA, V., CASELLI, M. C., CAPIRCI, O., AND PIZZUTO, E. Gesture and the emergence and development of language. In *Beyond Nature-Nurture: Essays in Honor of Elizabeth Bates*, M. Tomasello and D. I. Slobin, Eds. Lawrence Erlbaum Associates, 2004.

[18] WESTEYN, T., BRASHEAR, H., ATRASH, A., AND STARNER, T. Georgia tech gesture toolkit: Supporting experiments in gesture recognition. In *ICMI 2003: Fifth International Conference on Multimodal Interfaces* (2003), ACM Press.

Classifying Types of Gesture and Inferring Intent*

Chrystopher L. Nehaniv
*Adaptive Systems Research Group
School of Computer Science
Faculty of Engineering and Information Sciences
University of Hertfordshire
College Lane, Hatfield Herts AL10 9AB
United Kingdom
`C.L.Nehaniv@herts.ac.uk`

Abstract

In order to infer intent from gesture, a rudimentary classification of types of gestures into five main classes is introduced. The classification is intended as a basis for incorporating the understanding of gesture into human-robot interaction (HRI). Some requirements for the operational classification of gesture by a robot interacting with humans are also suggested.

1 Introduction: The Need for Classifying Gesture

The word *gesture* is used for many different phenomena involving human movement, especially of the hands and arms. Only some of these are interactive or communicative. The pragmatics of gesture and meaningful interaction are quite complex (cf. Kendon (1970); Mey (2001); Millikan (2004)), and an international journal *Gesture* now exists entirely devoted to the study of gesture. Applications of service or 'companion' robots that interact with humans, including naive ones, will increasingly require human-robot interaction (HRI) in which the robot can recognize *what* humans are doing and to a limited extent *why* they are doing it, so that the robot may act appropriately, e.g. either by assisting, or staying out of the way. Due to the situated embodied nature of such interactions and the non-human nature of robots, it is not possible to directly carry over methods from human-computer interaction (HCI) or rely entirely on insights from the psychology of human-human interaction. Insights from proxemics and kinesics, which study spatial and temporal aspects of human-human interaction (Hall, 1983; Condon and Ogston, 1967; Kendon, 1970) and some insights of HCI, e.g. recognizing the diversity of users and providing feedback acknowledgment with suitable response timing (e.g. (Shneiderman, 1998)), may also prove to be extremely valuable to HRI. Notwithstanding, the nascent field of HRI must develop its own methods particular to the challenges of embodied interaction between humans and robots. New design, validation, evaluation methods and principles particular to HRI must be developed to meet challenges such as *legibility*, making the robot's actions and behaviour understandable and predictable to a human, and '*robotiquette*', respecting human activities and situations (e.g. not interrupting a conversation between humans or disturbing a human who is concentrating or working intensely — without sufficient cause), as well respecting social spaces, and maintaining appropriate proximity and levels of attention in interaction. Part of meeting these challenges necessarily involves some understanding human activity at an appropriate level. This requires the capabilities of recognizing human gesture and movement, and inferring intent. The term "*intent*" is used in this paper in a *limited* way that refers to *particular motivation(s) of a human being that result in a gestural motion as relevant for human-robot interaction*.

In inferring the intent from a human's gesture it is helpful to have a classification of which type of gesture is being observed. Without a sufficiently broad classification, understanding of gesture will be too narrow to characterize what is happening and appropriate responses will not be possible in many cases.

While this paper does not attempt a comprehensive survey of the role and recognition of gesture in human-robot interaction, it does suggest inherent

---

[*] This paper is published in *Proc. AISB'05 Symposium on Robot Companions: Hard Problems and Open Challenges in Robot-Human Interaction*, The Society for the Study of Artificial Intelligence and Simulation of Behaviour, 2005.

limitations of approaches working with a too narrow notion of gesture, excluding entire classes of human gesture that should eventually be accessible to interactive robots able to function well in a human social environment.

The questions of how gestures are acquired and come to be recognized as meaningful by particular individuals in the course of their development (ontogeny of gesture and its recognition), and conventionalized, elaborated, or lost within particular cultures (evolution of gesture) are large and deep issues, but will not be addressed within the scope in this paper.

Knowing how to recognize and classify gesture may also serve to inform the design of robot behaviour, including gestures made by the robot to achieve legibility and convey aspects of the robot's state and plans to humans. This in turn will contribute to robot interaction with humans that is legible, natural, safe, and comfortable for the humans interacting with the robot.

2 Classification of Gestures

The following is a rough, tentative classification. Gestures are classed into five major types with some subtypes.

To begin to approach the complexity of gesture in the context of situated human-robot interaction, the rough classes of gesture described below are developed in order to provide a broad level of description and the first steps toward a pragmatic, operational definition that could be used by an autonomous system such as a robot to help it (1) to infer the intent of human interaction partners, and, as an eventual goal, (2) to help the robot use gestures itself (if possible) to increase the legibility of its behaviour.

Ambiguity of Gesture. It should be stressed that a single specific instance of a particular the kind of physical gestural motion could, depending on context and interaction history, reflect very different kinds of human intents. It will not always be possible to infer intent based solely on based the mechanical aspects of human movements (such as changes in joint angles) without taking context into account.

To approach this problem, a classification of gesture for inferring intent and assisting in the understanding of human activity should closely relate gesture with limited categories of intent in situated human activity. The classes of the tentative classification presented here thus correspond to and allow the (limited) attribution of intent on the part of humans. The classification is developed as an aid for helping robots to achieve limited recognition of situated human gestural motion so has to be able to respond appropriately if required, while these robots are working in an environment of ambient human activity (such as a home or office), in which, at times, the robots are also assisting or cooperating with the humans. *Applications of this classification will require the mapping of physical aspects of gestural motion in interactional contexts to the five gestural classes (and their subtypes) suggested here.*

2.1 Five Classes (with Subtypes)

1. **'Irrelevant'/Manipulative Gestures.** These include *irrelevant gestures*, *body / manipulator motion*, *side-effects of motor behaviour*, and *actions on objects*. Broadly characterized, manipulation by a human is here understood as *doing something to influence the non-animate environment or the human's relationship to it (such as position)*. Gestural motions in this class are manipulative actions (in this sense) and their side effects on body movement. These 'gestures' are neither communicative nor socially interactive, but instances and effects of human motion. They may be salient, but are not movements that are primarily employed to communicate or engage a partner in interaction. Cases include, e.g. motion of the arms and hands when walking; tapping of the fingers; playing with a paper clip; brushing hair away from one's face with one's hand; scratching; grasping a cup in order to drink its contents. (Note it may be very important to distinguish among the subtypes listed above for robot understanding of human behaviour.)

2. **Side Effect of Expressive Behaviour.** In communicating with others, motion of hands, arms and face (changes in their states) occur as part of the overall communicative behaviour, but without any specific interactive, communicative, symbolic, or referential roles (cf. classes 3-5)
   Example: persons talk excitedly raising and moving their hands in correlation with changes in voice prosody, rhythm, or emphasis of speech.

3. **Symbolic Gestures.** Gestural motion in symbol gesture is a *conventionalized signal in a communicative interaction.* It is generally a member of a limited, circumscribed set of gestural motions that have specific, prescribed interpretations. A symbolic gesture is used to trigger certain actions by a targeted perceiver, or to refer to some-

thing or substitute as for another signal according to a code or convention. Single symbolic gestures are analogous to discrete actions on an interface, such as clicking a button.
Examples: waving down a taxi for it to stop; use of a conventional hand signals (a command to halt indicated open flat hand; a military salute); nodding 'yes'; waving a greeting 'hello' or 'goodbye'.

Note that the *degree of arbitrariness* in such gestures may vary: The form of the gesture may be an arbitrary conventional sign (such as a holding up two fingers to mean 'peace', or the use of semaphores for alphabetic letters). On the other hand, a symbolic gesture may resemble to a lesser or greater extent iconically or, in ritualized form, a referent or activity.
Further examples: holding up two fingers to indicate 'two'; opening both (empty) hands by turning palms down to indicate a lack of something. Nearly all symbolic gestures are used to convey *content* in communicative interactions.

4. **Interactional Gestures.** These are gesture used to *regulate interaction with a partner*, i.e. used to initiate, maintain, invite, synchronize, organize or terminate a particular interactive, cooperative behaviour: raising a empty hand toward the partner to invite the partner to give an object; raising the hand containing an object toward the partner inviting them to take it; nodding the head indicating that one is listening. The emphasis of this category is neither reference nor communication but on gestures as mediators for cooperative action.[1] Interactional gestures thus concern regulating the form of interactions, including the possible regulation of communicative interactions but do not generally convey any of the content in communication. Interactional gestures are similar to class 1 manipulative gestures in the sense that they influence the environment, but in contrast to class 1, they influence the "animated environment" – *doing something to influence human agents (or other agents) in the environment*, but not by conveying symbolic or referential content.[2]

5. **Referential/Pointing Gestures.** These are *used to refer to or to indicate objects (or loci) of interest* – either physically present objects, persons, directions or locations the environment – by pointing (*deixis*[3] – showing), or indication of locations in space being used as proxies to represent absent referents in discourse.

Table 1 summarizes the five classes.

2.2 Target and Recipient of a Gesture

If a gesture is used interactively or communicatively (classes 2-5), it is important to recognize whether the gesture is directed toward the current interaction partner (if any) — which may the robot, another person (or animal) present in the context, or possibly neither (*target*). If pointing, what is the person pointing to? Who is the pointing designed to be seen by? (*recipient*). If speaking, to whom is the person speaking? If the gesture is targetted at or involves a contact with an object, this suggests it may belong to class 1 (or possibly 5, even without contact). A gesture of bringing an object conspicuously and not overly quickly toward an interaction partner is manipulative (in the sense explained in the discussion of class 1, since an object is being manipulated), but it may well at the same time also be a solicitation for the partner to take the object (class 4). Similarly if the partner has an object, an open hand conspicuously directed toward the partner or object may be a solicitation for the partner to give the object (class 4).

2.3 Multipurpose Gestures

It is possible for a single instance of a particular gesture to have aspects of more than one class or to lie intermediate between classes. As mentioned above,

---

[1] Note that we are using the word "cooperative" in a sense that treats regulating communication or interaction as an instance of cooperation.

[2] Some more subtle examples include putting one's hand on another person's arm to comfort them. Such actions, and others involving physical contact, may be quite complex to interpret as understanding them may require understanding and modeling the intent of one person to influence that state of mind of another. At this point, we class simply them with interactional gestures recognizing that future analysis may reveal deep issues of human-human interaction and levels of complexity beyond the rudimentary types of human intent considered here. A special case worthy of note is human contact with the robot, unless this is directly a manipulation of the robot's state via an interface - e.g. via button presses — which would fall into class 3 (symbolic gesture), non-accidental human contact with the robot is likely to be indicative of an intent to initiate or regulate interaction with the robot (class 4). Physical contact between humans might also involve expression of affection (kissing), or aggression (slapping, hitting) – which generally indicate types of human-human interaction it would be better for a robot to steer clear of!

[3] Deixis can involve a hand, finger, other directed motion, and/or eye gaze. Checking the eye gaze target of an interaction partner is commonly used to regulate reference and interaction; it develops and supports joint attention already in preverbal infants. Language, including deictic vocabulary (e.g. demonstratives such as the words "these" and "that"), and other interactional skills, typically develop on this scaffolding (see Kita (2003)).

handing over an object is both class 1 and 4. And, for example, holding up a yellow card in football has aspects of classes 1 and 3, object manipulation and conventional symbolic signal. Many ritualized symbolic gestures (class 3) also can be used to initiate or regulate interaction (class 4), e.g. the ‘come here’ gesture: with palm away from the recipient, moving the fingers together part way toward the palm; waving forearm and open hand with palm facing recipient to get attention. More complex combinations are possible, e.g. a gesture of grasping designed by the human to be seen by a recipient interaction partner and directed toward a heavy or awkwardly-sharped target object as a solicitation of the partner to cooperatively carry the object with the gesturer (classes 1, 4, 5).

2.4 Ritualization: Movement into Classes 3 and 4

Gestures that originate in class 1 as manipulations of the non-animate environment and the person’s relationship to it may become *ritualized* to invite interactions of certain types, e.g., cupping the hand next to the ear can indicate that person doing it cannot hear, so that the interaction partner should speak up. Originally cupping the hand near the ear served to improve a person’s ability to hear sounds in the environment from a particular direction (class 1), but it may be intended to be seen by a conversational partner who then speaks up (class 4). The hand cupped at the ear can even be used as a conventionalized symbol meaning ‘speak up’ (classs 3). Other examples of ritualization toward regulation of interaction and also symbolic gesture include mimicking with two hands the motions of writing on a pad as a signal to a waiter to ask for the bill; miming a zipping action across the mouth to indicate that someone should be ‘shut up’; or placing a raised index finger over lips which have been pre-formed as if to pronounce */sh/.*

2.5 Cultural and Individual Differences

Different cultures may differ in their use of the various types of gesture. Some symbolic gestures such as finger signs (e.g. the “OK” gesture with thumb and index finger forming a circle) can have radically different interpretations in other cultures, or no set interpretation depending on the culture of the recipient (e.g. crossing fingers as a sign of wishing for luck, or the Chinese finger signs for some numbers such as 6, 7, 8). Tilting the head back (Greece) or nodding the head (Bulgarian) are used symbolically for ‘no’, but would certainly not be interpreted that way in many other cultures. Cultures also differ in their types and scope of movement in (class 2) expressive gestures: Consider, for example, the differences of rhythm, prosody, hand motions, eye contact, and facial expressions accompanying speech between British, Italian, Japanese, and French speakers.

Within cultures, differences between different individuals’ uses of gestures can be regional, restricted to particular social groups within the culture, and vary in particularities (such as speed, repertoire, intensity of movement, etc.) between individuals according to preference or ontogeny. Elderly and young may employ gestures in different ways.

3 Some Related Work on Recognizing Gesture and Intent

The important role of gesture for intent communication in human-robot interaction is increasingly being acknowledged, although some approaches still focus only on static hand poses rather than dynamic use of more general types of gesture in context. A survey of hand gesture understanding in robotics appears in Miners (2002).

Multimodal and voice analysis can also help to infer intent via prosodic patterns, even when ignoring the content of speech. Robotic recognition of a small number of distinct prosodic patterns used by adults that communicate praise, prohibition, attention, and comfort to preverbal infants has been employed as feedback to the robot’s ‘affective’ state and behavioural expression, allowing for the emergence of interesting social interaction with humans (Breazeal and Aryananda, 2002). Hidden Markov Models (HMMs) have been used to classifying limited numbers of gestrual patterns (such as letter shapes) and also to generate trajectories by a humanoid robot matching those demonstrated by a human (Billard et al., 2004). Multimodal speech and gesture recognition using HMMs has been implemented for giving commands via pointing, one-, and two-handed gestural commands together with voice for intention extraction into a structured symbolic data stream for use in controlling and programming a vacuuming cleaning robot (Iba et al., 2002). Many more examples in robotics exist.

Most approaches use very limited, constrained, and specific task-related gestural repertoires of primitives, and do not attempt to identify gestural classes. They have tended to focus on a fixed symbolic set of gestures (possibly an extensible one, in which new gestures can be learned), or focus on only a few representatives from one or two of the gestural classes identified here (e.g. symbolic and pointing gestures).

Knowledge of specific conventional codes and signs can help the identification of particular signs within class 3, and also in determining that the gesture in fact belongs to class 3, i.e. is a symbolic communicative signal. Machine learning methods such as Hidden Markov Models may be used successfully to learn and classify gestures for a limited finite set of fixed gestures (e.g. (Westeyn et al., 2003)). It seems likely that HMM methods would be most successful with class 3 (symbolic gestures), but how successful they would be at differentiating between classes or within other classes remains uninvestigated at present.

4 Inferring the Intent of Gesture

Being able to classify gesture into one of the above classes gives us only a starting point for inferring the intent of the person making the gesture due to frequent ambiguity. Resolving this points to the important roles of context and interactional history. Thus, it is necessary to develop operational methods for recognizing the class of gesture in a particular context.[4]

To this it should help when

(a) the activity of the gesturer is known,

(b) previous and current interaction patterns are remembered to predict the likely current and next behaviour of the particular person,

(c) objects, humans and other animated agents in the environment are identified and tracked.

(d) the scenario and situational context are known (e.g. knowing whether a gesture occurs at a tea party or during a card game).

Knowing the above could help the robot classify the gesture and infer the intent of the human. Information on the state of human (e.g. working, thirsty, talking, ...) often can limit the possibilities.

4.1 Recognizing Intent from Gesture Given Interactional Context

If the interactional context of recent activity in which a gesture occurs is known, this can suggest possibilities for which classes (and subtypes) of gesture might be involved. Even giving data on the interactional context, including data on context, culture, individual differences, models of human activity and task aspects that relate to gesture, does not necessarily completely constrain the possible gesture nor its intent (if any). If the context suggests a particular identifying class (and subtype) for the gesture identified, this does not immediately lead to any certain knowledge of human intent behind it.

Data on the interaction history and context may help in determining the class of a gesture. If the class is known, then the set of possible gestures can remain large, or be narrowed significantly. Symbolic gestures (class 3) correspond to discrete symbols in a finite set, of which there may be only be a small number according to context or size of the given repertoire of the given symbolic gestural code. Interactional gestures (class 4) are likely to comprise a small, constrained class. Class 1 gestures are either “irrelevant”, or to be understood by seeking the intent of the associated motor action or object manipulation (e.g. grasping or throwing an object, arms moving as a side effect of walking). Class 5 (referential and pointing gestures) comprise a very limited class.

4.2 Typical Interactional Context of Gestures

A programme to apply the above classification can be developed as follows.

1. Identify the many, particular gestural motions that fit within each of the five classes. Some gestural motions will appear in more than one class. For example, the same mechanical motion of putting a hand and arm forward with the forearm horizontal and the hand open could indicate preparation to manipulate an object in front of the human (class 1), to show which object is being referred to (class 5), or to greet someone who is approaching, or to ask for an object to be handed over (both class 4).

2. Gestural motions identified as belonging to several classes need to be studied to determine in which contexts they occur: determining in which class(es) particular a instance of the gesture is being used may require consideration of objects and persons in the vicinity, the situational context, and the history of interaction.

3. Systematic characterizations of a physical gestural motion together with interactional contexts in which they are occur could then be used to determine the likely class.

---

[4] Knowledge of the immediate context in some cases needs to be augmented by taking into account of the broader *temporal horizon* of interactional history (cf. Nehaniv et al. (2002)).

Classification of gestural classes and associated (limited) categories of human intent		
Class	Name	Defining Characteristics and Associated Intent
1	‘Irrelevant’ or Manipulative Gestures	Influence on non-animate environment or human’s relationship to it; manipulation of objects, side effects of motor behavior, body motion
2	Side Effect of Expressive Behaviour	Expressive marking, (no specific direct interactive, symbolic, referential role) associated to communication or affective states of human
3	Symbolic Gestures	Conventionalized signal in communicative interaction; communicative of semantic content (language-like)
4	Interactional Gestures	Regulation of interaction with a partner; influence on human (or other animated) agents in environment but generally with lack of any symbolic/referential content used to initiate, maintain, regulate, synchronize, organize or or terminate various types of interaction
5	Referential/Pointing Gestures	Deixis; indicating objects, agents or (possibly proxy) loci of discourse topics, topics of interest; pointing of all kinds with all kinds of effectors (incl. eyes): referential, topicalizing, attention-directing

Table 1: **Five Classes of Gesture.** See text for explanation, details and examples. Note that some occurrences of the same physical gesture can be used in different classes depending on context and interactional history; moreover, some gestures are used in a manner that in the same instance belongs to several classes (see text for examples).

4.3 Updating the Interaction History

Attribution of intent related to gesture can then feedback into understanding of the situational context, including motivational state of the human performing the gesture, and becomes part of the updated interaction history, which can then help in inferring intent from ensuing gestures and activity.

5 Conclusions

In order to infer the intent of a human interaction partner, it may be useful to employ a classification of gesture according to some major types – five in the tentative classification proposed here – whose intent may be, in the five classes, absent / directed to objects or environment, incidentally expressive, symbolic, interactional, or deictic. A summary of the classes is given by Table 1.

In order to deploy the inference of intent on robots interacting with humans it will be necessary to operationalize the distinctions between these (sometimes overlapping) classes. This may require the use of knowledge of human activity, recognition of objects and persons in the environment, and previous interactions with particular humans, as well as knowledge of conventional human gestural referencing and expression, in addition to specialized signaling codes or symbolic systems.

The classification presented here suggests some requirements for the design and implementation of systems inferring intent from gesture based on this classification. These requirements might be realized in a variety of different ways using, e.g. continuous low-key tracking or more detailed analysis, event-based and/or scenario-based recognition, and prediction of human activity based on models of human activity flows (with or without recognition of particular humans and their previous interactions), depending the particular needs of the given human-robot interaction design and the constraints and specificity of its intended operational context. Design of a robot restricted to helping always the same user in the kitchen environment would be quite different from one that should be a more general purpose servant or companion in a home environment containing several adults, children and pets, but the classification presented here is applicable in informing the design of gesture recognition for inferring intent in either type of system, and for designing other HRI systems.

Finally, effective human-robot interaction will require generation of gestures and feedback signals by the robot. The classification given here can suggest categories of robotic gestures that could be implemented to improve the *legibility* to humans of the robot's behaviour, so that they will be better able to understand and predict the robot's activity when interacting with it.

Acknowledgments

The work described in this paper was conducted within the EU Integrated Project COGNIRON ("The Cognitive Robot Companion") and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020.

The classification presented here is developed by the author in response to discussions in the COGNIRON project, especially with Rachid Alami, Kerstin Dautenhahn, Jens Kubacki, Martin Haegele, and Christopher Parlitz. Thanks to Kerstin Dautenhahn for useful comments on an earlier draft of this paper.

This paper extends and supercedes University of Hertfordshire School of Computer Science Technical Report 419 (December 2004).

References

A. Billard, Y. Epars, S. Calinon, G. Cheng, and S. Schaal. Discovering optimal imitation strategies. *Robotics & Autonomous Systems,* Special Issue: Robot Learning from Demonstration, 47(2-3): 69–77, 2004.

Cynthia Breazeal and Lijin Aryananda. Recognition of affective communicative intent in robot-directed speech. *Autonomous Robots*, 12(1):83–104, 2002.

W. S. Condon and W. D. Ogston. A segmentation of behavior. *Journal of Psychiatric Research*, 5:221–235, 1967.

Gesture. ISSN: 1568-1475. John Benjamins Publishing Co., The Netherlands `http://www.benjamins.com/cgi-bin/t_seriesview.cgi?series=GEST`.

Edward T. Hall. *The Dance of Life: The Other Dimension of Time*. Anchor Books, 1983.

Soshi Iba, Christian J. J. Paredis, and Pradeep K. Khosla. Interactive multi-modal robot programming. In *Proceedings of the 2002 IEEE International Conference on Robotics and Automation, Washington D.C., May 11-15, 2002*, 2002.

Adam Kendon. Movement coordination in social interaction: Some examples described. *Acta Psychologica*, 32:100–125, 1970.

Sotaro Kita, editor. *Pointing: Where Language, Culture and Cognition Meet*. Lawrence Erlbaum Associates, Inc, 2003.

Jacob Mey. *Pragmatics: An Introduction*. Blackwell Publishers, 2001.

Ruth Garrett Millikan. *The Varieties of Meaning: The 2002 Jean Nicod Lectures*. MIT Press/Bradford Books, 2004.

William Ben Miners. *Hand Gesture for Interactive Service Robots*. MSc Thesis, The University of Guelph, Faculty of Graduate Studies, August 2002.

Chrystopher L. Nehaniv, Daniel Polani, Kerstin Dautenhahn, René te Boekhorst, and Lola Cañamero. Meaningful information, sensor evolution, and the temporal horizon of embodied organisms. In *Artificial Life VIII*, pages 345–349. MIT Press, 2002.

Ben Shneiderman. *Designing the User Interface: Strategies of Effective Human-Computer Interaction*. Addison-Wesley, 3rd edition, 1998.

Tracy Westeyn, Helene Brashear, Amin Atrash, and Thad Starner. Georgia tech gesture toolkit: Supporting experiments in gesture recognition. In *ICMI 2003: Fifth International Conference on Multimodal Interfaces*. ACM Press, 2003.

Navigation in the Presence of Humans*

E. A. Sisbot, R. Alami and T. Simeon
Robotics and Artificial Intelligence Group
LAAS/CNRS
7, Avenue du Colonel Roche, 31077 Toulouse cedex 4
France
{*easisbot, rachid, nic*}*@laas.fr*

K Dautenhahn, M Walters, S Woods,
K L Koay and C Nehaniv
University of Hertfordshire
College Lane, Hatfield Herts AL10 9AB
United Kingdom
K.Dautenhahn@herts.ac.uk

***Abstract*— Robot navigation in the presence of humans raises new issues for motion planning and control since the humans safety and comfort must be taken explicitly into account.**

We claim that a human-aware motion planner must not only elaborate safe robot paths, but also plan good, socially acceptable and legible paths. Our aim is to build a planner that takes explicitly into account the human partner by reasoning about his accessibility, his vision field and potential shared motions.

This paper focuses on a navigation planner that takes into account the humans existence explicitly. This planner is part of a human-aware motion and manipulation planning and control system that we aim to develop in order to achieve motion and manipulation tasks in a collaborative way with the human.

We are conducting research in a multidisciplinary perspective, (1) running user studies and (2) developing an algorithmic framework able to integrate knowledge acquired through the trials. We illustrate here a first step by implementing a human-friendly approach motion by the robot.

I. INTRODUCTION

The presence of humans in the robot environment and the necessity to interact with them raise a number of new questions and challenges. Clearly, the human should be taken explicitly into account in all steps of the robot design.

This paper addresses issues related to the close interaction between humans and robots from the standpoint of the motion decisions that must be taken by the robot in order to ensure a:

- A safe interaction, i.e., that cannot harm the human,
- A reliable interaction, i.e, that achieves the task adequately considering the motion capacities of the robot, and
- A user friendly interaction, i.e, that takes into account a motion model of the human as well as his preferences and needs.

Let us consider a "simple fetch and carry task" as illustrated in figure 1 for a socially interactive robot [8]. The robot has to perform motion and manipulation actions and should be able to determine where a given task should be achieved, how to place itself relatively to a human, how to approach him, how to hand an object.

Our goal is to develop a robot that is able to take into account "social constraints" and to synthesise plans compatible with human preferences, acceptable by humans and easily legible in terms of intention.

Fig. 1. A "fetch-and-carry" Scenario

All these questions have a particular flavour when the robot is a humanoid. Indeed, by its shape and functional abilities, it can act in very close interaction with humans, adopt postures and perform motions that are easily understandable by its human partners.

In this paper, we concentrate more precisely on navigation in the vicinity of humans. We are conducting research in a multidisciplinary perspective, (1) running user studies and (2) developing an algorithmic framework able to integrate knowledge acquired through the trials. We illustrate here a first step by implementing a human-friendly approach motion by the robot. Section II discusses related work. Section III presents trials that we have conducted in order to find out about subject preferences for the robot approach directions. Section IV presents the main characteristics of our navigation planner. Finally, we illustrate the outputs of a prototype implementation in section V.

---

* The work described in this paper was partially conducted within the EU Integrated Project COGNIRON ("The Cognitive Companion") and funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020.

II. HUMAN-ROBOT INTERACTION IN LITERATURE

A key issue is safety problems, where humans exist nearby robots. In industrial robotics, safety is assured by not allowing humans in a certain perimeter around robots and having emergency stop buttons [13]. Actually there is no interaction in these cases.

In a recent work by Nonaka et al. [18], the concept of safety has been studied by two aspects: ”physical” safety and ”mental” safety of human. Physical safety means that the robot do not physically injure humans. Mental safety, on the other hand, means that the motions of the robot do not cause any unpleasantness like fear, shock, surprise to human.

The physical safety is an absolute need for the human-robot interaction. It must be assured at the hardware and software design process of the robot. We can classify the safety strategies into two different types [11]: design strategies and control strategies. Besides new designs [3], [24] that will ensure safety at the physical level, fault-tolerant approaches [16] tend to detect and limit the consequences of hardware and software problems. A danger criterion is generally considered in control strategies and robot motions are executed by minimising this criterion [14].

With these approaches physical safety is assured by avoiding collision with human and by minimising the intensity of the impact in case of a collision. Another direction towards the motion in presence of humans is the research made for smart wheelchairs. Although there is not a real interaction between chair and human in a direct sense, the wheelchair motion needs to take into account the humans comfort [19].

In usual interactions between humans, some non written rules are respected that determine the distance between two persons (see the proxemic theory of E. T. Hall [10]). The robot should comply to similar conventions [4].

Other works try to imitate human motions for a better understanding of how humans behave in social environments. A recent work [2] makes robot place himself like humans in a conversation. We must note that this behaviour is only imitating humans self-placement.

Another approach that deals not only with safety but also implicitly comfort issues is the work on velocity profiles along a planned trajectory made by Alami et al. [1] where a robot adapts its trajectory and its speed to optimise the execution time and also to guarantee that no collision will occur. Although the human is not considered explicitly, this method guarantees a motion without collision by taking into account the sensor capabilities of the robot. Since the sensors have a certain range, it is likely necessary to slow down in some places of the robot's trajectory where the sensors are blocked by narrow passages or corners. And a velocity profile is found by optimising the execution time.

While motion planning and control for humanoid robots deal with their specificities in terms of motion and manipulation [12], we claim that explicit reasoning on humans should also be integrated.

Although several authors propose motion planning or reactive schemes considering humans, there is no contribution that tackles globally the problem as we propose to do.

III. HUMAN-ROBOT APPROACH DIRECTION TRIALS

This section presents some relevant results from a demonstration Human-Robot Interaction trial event, which was run at the AISB Convention at the University of Hertfordshire (UH). The UH team are primarily interested in the human perspective of how robots could be useful in domestic environments; in particular the roles, tasks, and social behaviour that will be necessary for robots to exhibit in order to integrate into normal domestic situations.

In order to study human-robot relationships, they typically run HRI trials using carefully devised test scenarios, where the human centred view can be collected using a variety of methods. The “Wizard of Oz” (WOZ) technique, where robots are remotely controlled by human operators, is widely used in HRI studies where the human reactions to robot behaviour are investigated; see [22], [9], [17], [20]. The WOZ technique can quickly test proposed complex robot behaviours or capabilities. Subjects' opinions of social acceptability towards robot behaviour can be assessed before committing resources to develop fully autonomous capabilities, which may be expensive, technically difficult or even impossible to implement to date.

A. The HRI Trial Method

The chosen scenario involved a robot using different approach directions to bring a seated subject an object. The aims of the trial were to find out about subject preferences for the robot approach directions.

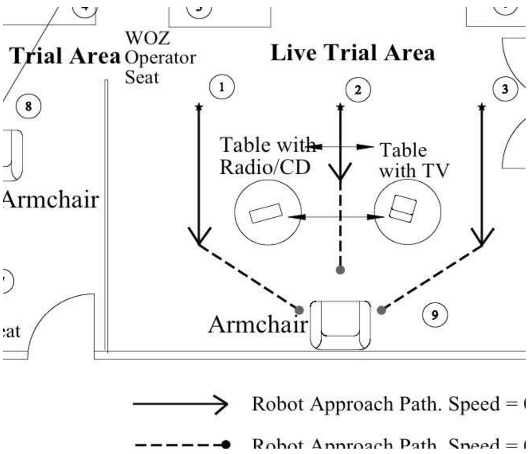


Fig. 2. Diagram of HRI approach trial experiment areas

The AISB demonstration event had an audience and was performed under non-laboratory conditions using 38 volunteers from the convention participants. There was also a follow up study carried out under more controlled conditions with 15 subjects and one of the aims of this trial was to verify and check results obtained from the demonstration study. The trial area was identically laid out for both trials to resemble a simulated living room with a chair and two tables. The subject was seated in the chair. To the left front and right front of the chair, two tables were arranged (with room for the robot to pass by) in front of the chair. One of the tables had a television placed upon it; the other had a radio and CD player. Posters were attached to the wall directly opposite the chair to provide a more comfortable atmosphere for the subject. The robot was driven to the appropriate start position by the WOZ operator. The robot's approaches to the subject were fully autonomous. The WOZ operator was seated at a table in the far corner of the room. Subjects were told that the robot would be controlled by the operator while it was driven to the three start positions, but would be approaching them autonomously to bring them the TV remote control. This was reinforced as the WOZ operator made notes and did not press any of the robot control keys (on the robot control laptop) while it approached the subject. (see Fig. 3 for example video clips).

B. The HRI Trial Scenario

The context was that the subject had arrived home, tired after a long day at work and rested in an armchair (Fig. 2). After looking around for the TV remote control, the subject then asked the robot to fetch it for them as they were too tired to get up. The robot then brought the remote control to the subject. It was explained to the subject that the robot was new to the household and it was necessary to find out which approach direction the subject preferred; either from the front (2), the left (1) or the right (3). The three possible paths taken by the robot are shown in Fig. 2. In order to justify the scenario of the robot fetching the remote control, one of the tables had a (switched off) TV set upon it. The other table had a CD-Radio unit. Our expectations prior to the trials were that subjects would prefer the approach from the front, since the robot was then fully visible at all times.

Fig. 3. Clips from the robot to human approach trials

A short introductory questionnaire was used to gain basic demographic and personal details from the subjects. At the end of each HRI trial a short questionnaire was used to assess the subjects' views on approach direction, approach speed, stopping distances, comfort levels and practicality for the different approach directions. The subjects' reactions to both live and video based HRI trials were also recorded by a single tripod mounted camera placed at an appropriate point, either (4), (5) or (6) (Fig. 2). Twenty one males (54%) and 18 females (46%) participated in the AISB robot approach direction trials. The mean age of subjects was 36 years (range: 22-58). Thirty five subjects (95%) of subjects were right handed and 2 subjects (5%) were left handed.

C. Demonstration Trial Results

a) Approach Direction Preferences: Fig. 4 illustrates that 60% (N: 23) of subjects stating preferred the right robot approach direction, followed by 24% (N: 9) preferring the left approach and just 16% (N: 6) preferring the front approach. An overriding majority of subjects stated least preferring the front robot approach direction (N: 31, 80%). Few subjects least preferred the left and right approach directions.

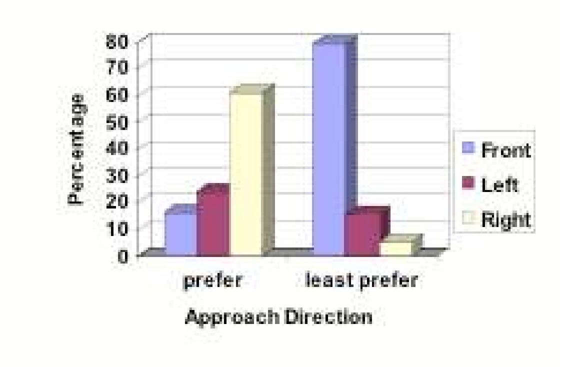


Fig. 4. Approach Directions Most and Least Preferred

Chi-square cross-tabulations revealed a non-significant association between gender and the robot approach direction preferred (X2 (2, 38) = 3.77, p = 0.1). More females stated that they preferred the robot front approach direction compared to males, and more males preferred the right robot approach direction compared to females (See Fig. 5).

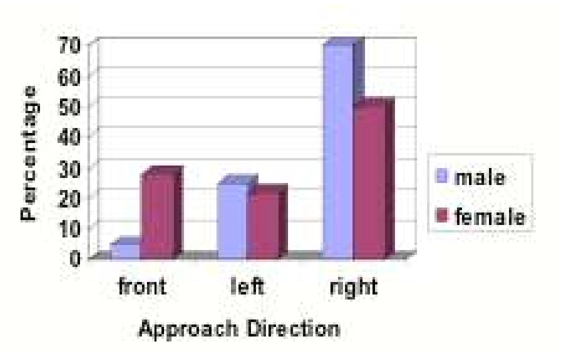


Fig. 5. Approach Direction Preferences by Gender

A significant relationship was found between gender and least preferred robot approach direction (X2 (2, 39) = 7.09, p = 0.03). Significantly more males stated least preferring the front robot approach direction compared to females (males: 95%, females: 61%). More females stated least preferring the right robot approach direction compared to males (males: 0%, females: 11%). Chi-square cross-tabulations revealed no significant relationships between age, handedness and approach directions preferred and least preferred.

b) Approach Distance Preferences: Subjects were asked to provide an overall rating (for all 3 approach directions) for the robot approach distance. 76% (N: 28) of subjects stated that the distance between them and the robot was 'about right', followed by 19% (N: 7) who felt that the robot was to 'too far' from them. Only 5% (N: 2) of subjects stated that the robot approached them too closely.

D. *HRI Trial Conclusions*

These results indicate that a large majority of subjects disliked robot approaching from the front approach direction. Some subjects sat with their legs over the side of the chair (see Fig. 2) and for these few cases, it may be that in this case a frontal approach (relative to the chair) was preferred. Thus there may be a more general rule that people do not like the robot to approach from the direction which their legs poke out. In the case of subjects who rated the front stopping distance as being too far, we observed that these subjects usually had their legs stretched out in front of them. This caused the robot to stop (due to the robot's stopping safety mechanism) when it reached the subject's feet, rather than moving close enough for them to reach the TV remote control comfortably.

IV. A NAVIGATION PLANNER

Today, the classical motion planning methods [15] are quite efficient to find feasible paths. However, the presence of humans in the environment drastically changes the notion of acceptable paths. In a human-robot interaction context, the computed paths do not only need to be collision-free but must also take into account the human comfort. This is illustrated on figure 6 which shows two paths possibly produced by a classical motion planner. Obviously, both paths are badly chosen since one path passes too close to the wall, causing a surprise to human, and the other passes behind of the human also causing some discomfort.

The User studies with humans and robots reported in the previous section (see also [23][2]) provide a number of properties required when dealing with humans. Only very limited works consider such comfort and legibility issues, often in an ad hoc manner. We describe below a new technique that allows to integrate such additionnal constraints in a more generic way. First, we introduce three criteria to the motion planning stage to ensure the safety and comfort. The robot must take into account these three criteria at the planning stage along with the more common aspects of path planning such as obstacle avoidance and shortest path finding.

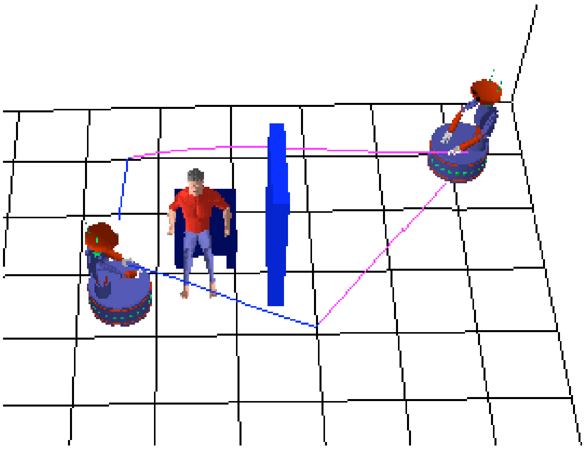


Fig. 6. Two paths found by classical motion planning

Each criterion is represented by a numerical potential stored in a 2D grid combining various costs. These costs are highly related to the humans' state, capabilities and preferences. The grid G can be defined as follows:

$$G = (M_{n,p}, H_1 \ldots H_n)$$

where $M_{n,p}$ is a matrix containing $n * p$ cells represented by ,$a_{x,y}$, the cost of the coordinate (i, j) in a 2D plane and $H_1 \ldots H_n$ is a list of humans in the environment. A human H_i is modeled by $H_i = (St, State_1 \ldots State_n)$ where St is the structure and kinematics of the human and $State_i$ is a human state defined by a number of cost parameters and state description:

$$State_i = (Name, Conf, Param)$$

where $Name$ is the name of the state (for ex. $Name = SITTING, STANDING$), $Conf$ is the humans configuration in that state and $Param$ represents the data needed to compute costs according to that state.

From the user studies above, we extracted three criteria that will allow the robot to be more human friendly.

A. *Safety Criterion*

The first criterion, called safety criterion, mainly focuses on ensuring the safety by controlling the distance between robot and human. The robot, if possible, must avoid approaching too much to human, and in some cases a certain perimeter around human must not be allowed to pass through. However, the robot must be able to approach the human because of the necessity of their interaction (for example to handle some object to the human). Hence, this distance between the robot and the human is not uniform and fixed, but depends on the interaction. The feeling of security is highly dependent to the humans personality and physical capabilities. For example, a robotics scientist feels much comfortable when a robot is around than an other person who sees the robot first time. Also, an elderly person can feel safer when the robot is 4 meters away from him than a curious teenager who feels safe even when the robot approaches very close. The humans current state plays also an important role as the safety feeling differs highly when the human is sitting than when he is standing up. When the human is sitting, as his mobility is reduced, he tends to have a low tolerance to the robot getting close. On the contrary when standing up he gets a higher mobility, therefore allowing the robot to come closer.

The user studies on the spatial interaction between people indicate that the spaces 1-3 meters away from humans are considered as interaction spaces with non-friends and $> 3m$ zones are considered as public zones. These studies can be used to define the distance limitations between robots and humans [10], [5], along with the studies of social spaces between robots and humans [23]

The safety grid contains a human centered gaussian form of cost distribution. Each coordinate (x, y) in this grid contains a cost inversely proportional to the distance to the human. Then, when the distance between the human and a point in the environment (in the grid) $D((x_i, y_j))$ is greater than the distance of another point $D((x_k, y_l))$, we have $Cost(x_k, y_l) > Cost(x_i, y_j)$. Since the safety concerns loose their importance when the robot is far away from the human, the cost also decreases when getting farther from the human, until some maximal distance at which it becomes null.

Figure 7 shows a computed safety grid attached to a human who is sitting on a chair. The vertical lines represent the cost associated to each cell. As shown by the figure, the cost is maximal at the human position and the cost range considered for a sitting human is approximately 3 meters [1].

As mentionned above, this cost highly depends on the human state and figure 8 shows a different cost computed for a standing human. As his mobility increases, the cost influence range shrinks to approximately 2 meters .

Other types of human state can be easily taken into account. For example, costs associated to states like sleeping, awake, ... can be easily handled with corresponding $Params$.

Once the grid is computed, searching for a minimum cost path will avoid to move too close to the human since approaching the human is more costly than staying far away.

B. Visibility Criterion

The human comfort is another criterion to be considered when dealing with human-robot interactions. In particular, the human generally feels more comfortable when he sees the robot. Therefore, we introduce another criterion, called visibility criterion, in order to help the robot to stay in the field of view of the human during its motions.

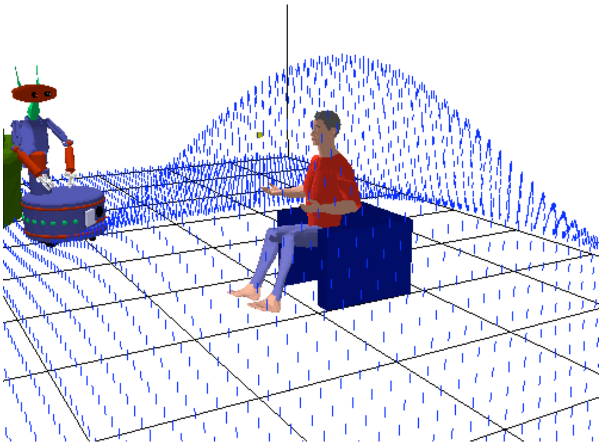


Fig. 7. Safety grid when human is sitting

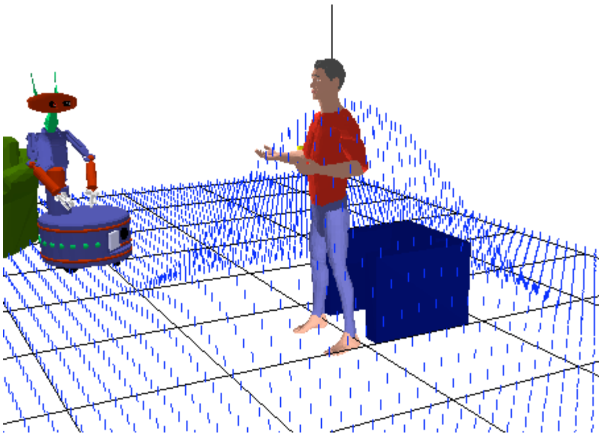


Fig. 8. Safety grid when human is standing

The visibility grid is constructed according to costs reflecting the effort required by the human to get the robot in his field of view. Grid points located in a direction for which the human only has to move his eyes have a lower cost than positions requiring to move the head in order to get the robot in the field of view. Also, when the robot is far away from the human, the effect of the visibility must decrease. The computed visibility costs are shown on figure 9. The zone situated in front of the human has very low costs. On the contrary, the zone situated behind the human has higher costs. As the visibility has a certain range of effect, the intensity decreases according to the distance to the human and its effect becomes negligible after 3-4 meters. Also, since the grid is attached to the head of the human, the computed costs are actualized when the human changes his field of view (turn his head) in planning and/or execution stage.

C. Hidden Zones

In the grids illustrated above, the costs are calculated without taking into account the obstacles in the environment. However,

---

[1] These values are only estimates and can be changed according to the context.

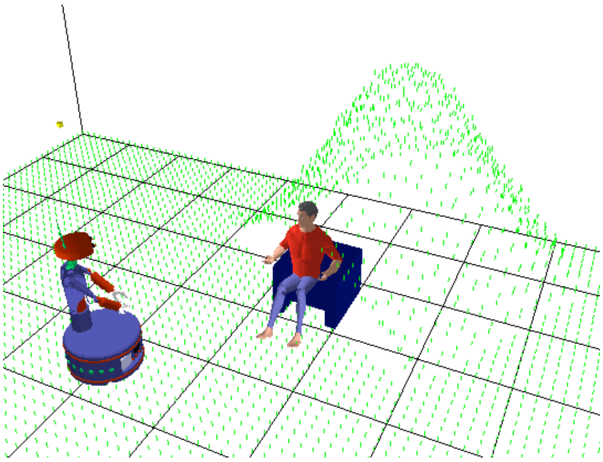


Fig. 9. Visibility grid

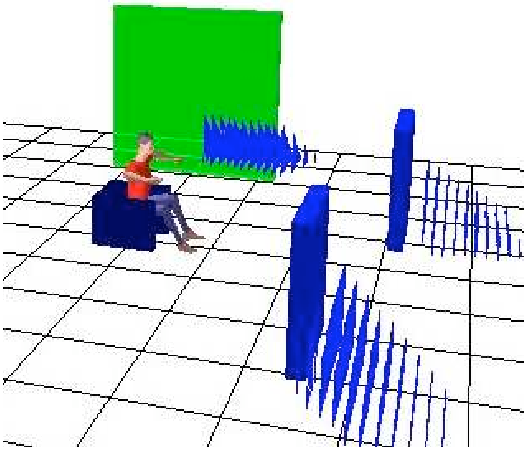


Fig. 10. Decreasing costs of hidden zones grid

obstacles in the close vicinity of the human can have various effects on the safety and comfort. If the robot is behind an obstacle, the human would feel much comfortable because the obstacle would block the direct way between human and the robot. So the distance criterion must be cancelled in the zones located behind the obstacles.

On the other hand, as the robot becomes hidden when it passes behind an obstacle, the visibility costs do not correspond anymore to physical realities. To handle this issue, we introduce another criterion additional to visibility and safety, called "hidden zones" criterion. This criterion helps to determine better costs for positions hidden by the obstacles.

Hence, an important effect of obstacles to the comfort of the human is the surprise factor. When the robot is hidden by an obstacle close to the human and suddenly appears in the human field of view, it can cause surprise and fear. To avoid this effect, we must discourage the robot to pass behind an obstacle too closely, and must allow it to get into the humans field of view when sufficiently far from the human. This can be done by putting costs to the zones hidden from the view by the obstacles.

The costs in the hidden zone grid is inversely proportional to the distance between the human and the robot. The range of the effect of the surprise factor is approximately 3m, so the costs decrease to zero in the 3m perimeter and remains null for the other grid points (Fig. 10).

D. Path planner

Once the safety, visibility and hidden zones grids have been computed, they are merged to one single grid that the robot will search for a minimum cost path. Note that we do not compute the four grids (3 criteria + 1 final) explicitly but just the costs of the grid cells necessary during search. Different ways can be used to merge the grid costs. A first way can be to compute the overal cost from the ponderated sum of the elementary costs

$$Cost_{merged}(x,y) = w_1 Cost_{safety}(x,y) + w_2 Cost_{visibility}(x,y)$$

where (x,y) is a grid point, w_1 is the weight of the safety grid and w_2 is the weight of the visibility grid.

Another way can be to consider the maximum cost values when merging the grids

$$Cost_{merged}(x,y) = max(Cost_{safety}(x,y), Cost_{visibility}(x,y))$$

Note that we do not merge hidden zones grid with the other 2 grids. That is mainly because hidden zones grids serves as a replacement of this 2 grids for positions where the robot could be seen if it wasn't blocked by an obstacle. The final grid is computed by:

```
 if ($R$ is on $(x,y)$ AND
$R$ is in field of view of $H_i$ AND
$H_i$ cannot see $R$ because of an obstacle $O$)
then $Cost_{final}(x,y)$ <- $w_3 Cost_{hiddenzones}(x,y)$
else $Cost_{final}(x,y)$ <- $Cost_{merged}(x,y)$
```

Our planner can use both ways depending on the task and on the balance between criteria. For example, for an urgent task, the importance of the visibility grid is less than the safety grid so that the robot does not take to much into account the visibility.

Once the final grid is computed, the cells corresponding to the obstacles in the environment are labeled as forbidden and an A* search is performed to find minimum-cost path between given two positions of the robot. The computed path is collision-free and also takes into account the human comfort and safety.

V. RESULTS

The navigation planner is implemented[2] within the Move3D [21] software platform developed at LAAS.

---

[2] A number of mpeg animations of computed motions can be found at http://www.laas.fr/ easisbot

Figure 11 shows a first path computed in the presence of two humans looking each other. One can note that the computed solution corresponds to a direct path remaining at equal distance from both humans. Consider now the same problem solved in a situation where one of both humans is turned back. Even the previous path is feasible, the planner chooses a more confortable path along which the robot remains visible for both humans (figure 12).

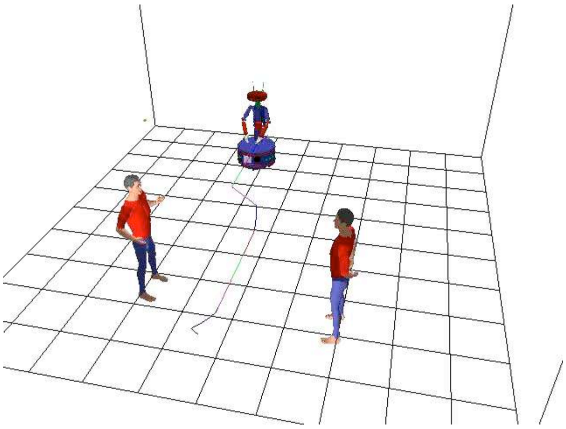


Fig. 11. The robot chooses a path with maximum safety and comfort

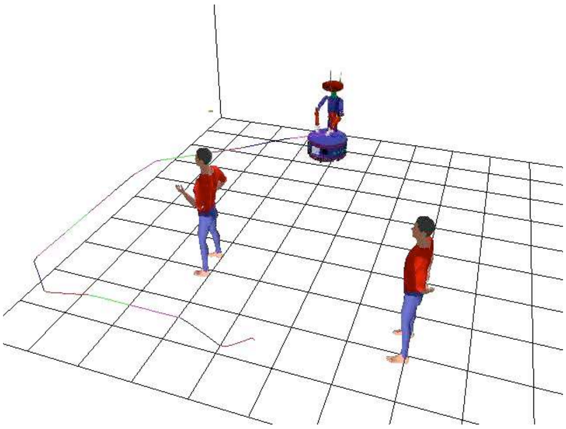


Fig. 12. The path changes according to the visibility grids

Figure 13 illustrates another scenario with two humans sitting in a room. The robot is initially located in the right corner of the room and has to move next to the human hidden by the wall obstacle. The figure shows the safety and visibility grids computed for each human, and the hidden zone created by the wall.

The minimum cost path computed by the planner (figure 14) has the following characteristics:

- The robot does not approach too close to the both of the humans. It chooses a solution that only enters in the humans 3m zone in the last portion of the path.
- The robot remains as visible as possible along the path. Because of the hidden start position, there is no possibility to be in the human field of view at the beginning of path. Therefore the planner chooses to pass behind the wall instead of passing behind the human.
- The robot is not to close to the human when it appears in its field of view. The transition from the invisible zone behind the wall to the visible one is sufficiently far from the human to avoid any surprise effect. Then the robot can approach to the human to reach its final position.

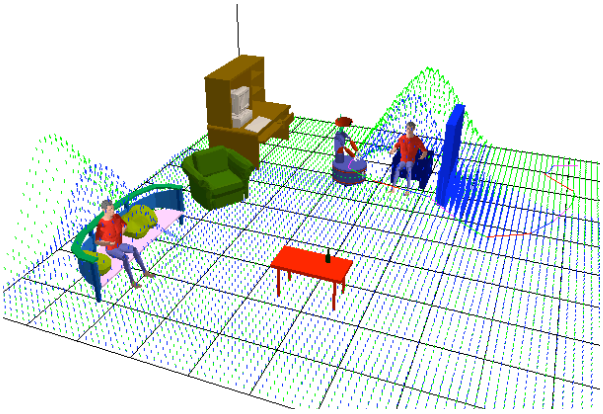


Fig. 13. The 3 types of grids

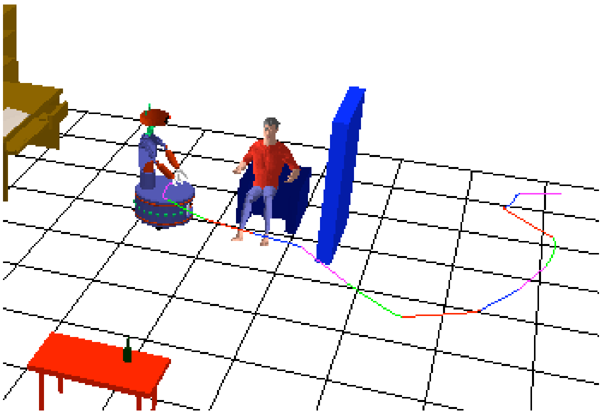


Fig. 14. The computed path takes into account the safety and the comfort of the human

As in real world scenarios there can be multiple objects and humans in motion, if the environment changes during execution, the calculated path can loose its validity. This trajectory can be replaced very fastly by replanning that unvalid trajectory. As HAMP's planning method is fast (the trajectories for both exaples are produced in less than a second with and AMD Athlon 1.8 Mobile processor), replaning allows us to be sufficiently reactive to the environment changes and to execute trajectories smoothly making the replanning totaly transparent to the humans. .

VI. CONCLUSION AND FUTURE WORK

We have proposed a navigation planner that takes into account the humans existence explicitly and that not only elaborates safe robot paths, but also plans good, socially acceptable and legible paths.

It is based on a multidisciplinary perspective. The algorithmic framework that we propose is able to integrate results provided by user studies.

This is a small step towards a very ambitious goal. Indeed, there a already a number of extensions that can be envisaged. For instance, it is clearly necessary to consider speed and accceleration and their influence on human comfort and acceptability of the robot actions. Besides, it remains to implement such schemes on real robot and to conduct validation tests.

We are also planning to develop a manipulation planner in order to allow the robot to hand objects to a human while respecting the safety and social constraints.

REFERENCES

[1] R. Alami, T. Simeon and K. Madhava Krishna, "On the influence of sensor capacities and environment dynamics onto collision-free motion plans", Proc. IEEE Int. Conf. on Intelligent Robots and Systems, 2002.

[2] P. Althaus, H. Ishiguro, T. Kanda, T. Miyashita and H. I. Christensen, "Navigation for human-robot interaction tasks", Proc. in IEEE Int. Conf. on Robotics & Automation, New Orleans, USA, 2004.

[3] A. Bicchi and G. Tonietti, "Fast and Soft Arm Tactics: Dealing with the Safety-Performance Trade-Off in Robot Arms Design and Control", *IEEE Robotics and Automation Magazine*, IEEE Volume 11, Issue 2, June 2004 Page(s):22 - 33

[4] R. Chatila , R. Alami , T. Simeon , J. Pettre , L. Jaillet, "Safe, reliable and friendly interaction between humans and humanoids", Rapport LAAS No02530, *3rd 3rd IARP International Workshop on Humanoid and Human Friendly Robotics*, Tsukuba (Japan), 11-12 December 2002, pp.83-87

[5] H. I. Christensen and E Pacchierotti, "Embodied social interaction for robots", Proc. AISB'05 Symposium on Robot Companions: Hard Problems and Open Challenges in Human-Robot Interaction, 2005, pp.40-45.

[6] A. Clodic, V. Montreuil, R. Alami, R. Chatila, "A Decisional Framework for Autonomous Robots Interacting with Humans" *Proc. IEEE Ro-man 2005*, Nashville, August 2005.

[7] Cogniron Project Web Site, "http://www.cogniron.org"

[8] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," Robotics and Autonomous Systems, vol. 42, pp. 143-166.

[9] A. Green, H. Huttenrauch and K Severinson Eklundh, "Applying the Wizard of Oz Framework to Cooperative Service discovery and Configuration," Proc. IEEE Ro-man 2004, 13th IEEE International Workshop on Robot and Human Interactive, Oka-yama, Japan, IEEE Press, 2004.

[10] E. T. Hall, "The Hidden Dimension", Doubleday, Garden City, N.Y., 1966.

[11] K. Ikuta, H. Ishii and M. Nokata, "Safety evaluation methods of design and control for human-care robots", The International Journal of Robotics Research, Vol 22, No. 5, 2003, pp. 281-297.

[12] J.J. Kuffner, K. Nishiwaki, S. Kagami, M. Inaba, and H. Inoue. "Motion planning for humanoid robots". In *Proc. 20th Int'l Symp. Robotics Research (ISRR'03)*, Italy, October 2003.

[13] D. Kulic and E. Croft, "Strategies for safety in human-robot interaction", Proc. IEEE Int. Conf. on Advanced Robotics, 2003, pp. 810-815.

[14] D. Kulic and E. Croft, "Safe planning for human-robot interaction," Proc. in IEEE Int. Conf. on Robotics & Automation, New Orleans, USA, 2004.

[15] J.C. Latombe, Robot Motion Planning, Kluwer Academic Publishers, Boston, MA, 1991.

[16] B. Lussier, A. Lampe, R. Chatila, J. Guiochet, F. Ingrand, M.O. Killijian, D. Powell, "Fault tolerance in autonomous systems: how and how much?", *4th IARP/IEEE-RAS/EURON Workshop on Technical Challenges for Dependable Robots in Human Environments*, Nagoya (Japan), 16-18 Juin 2005, 10p.

[17] D. Maulsby, S. Greenberg and R Mander, "Prototyping an intelligent agent through Wizard of Oz," Proc. ACM SIGCHI Conference on Human Factors in Computing Systems, Amsterdam, The Netherlands, ACM Press, pp. 277-284, 1993.

[18] S. Nonaka, K. Inoue, T. Arai and Y. Mae, "Evaluation of human sense of security for coexisting robots using virtual reality", in. Proc IEEE Int. Conf. on Robotics & Automation, New Orleans, USA, 2004.

[19] R. S. Rao, K. Conn, S. H. Jung, J. Katupitiya, T. Kientz, V. Kumar, J. Ostrowski, S. Patel and C. J. Taylor, "Human-robot interaction: Application to smart wheelchairs," Proc. in IEEE Int. Conf. on Robotics & Automation, Washington, DC, 2002.

[20] B. Robins, K. Dautenhahn and J. Dubowski, "Investigating Autistic Children's Attitudes Towards Strangers with the Theatrical Robot - A New Experimental Paradigm in Human-Robot Interaction Studies," Proc. 13th IEEE International Workshop on Robot and Human Interactive. Oka-yama, Japan, IEEE Press, pp. 557-562, 2004.

[21] T. Siméon, JP. Laumond, F. Lamiraux. Move3D: a generic platform for motion planning. *4th International Symposium on Assembly and Task Planning*, Japan, 2001.

[22] M. Walters, K. L. Koay and K. Dautenhahn, "Practical and Methodological Challenges in Designing and Conducting Interaction Studies with Human Subjects," Robot Companion Symposium at the. Convention on Social Intelligence and Interaction in Animals, Robots and Agents. (AISB'05). 12th - 15th April 2005. University of Hertfordshire, UK.

[23] M.L. Walters, K. Dautenhahn, R. te Boekhorst, K. L. Koay, C. Kaouri, S. Woods, C. Nehaniv, D. Lee, I. Werry , "The Influence of Subjects Personality Traits on Personal Spatial Zones in a Human-Robot Interaction Experiment", *Proc. IEEE Ro-man 2005*, Nashville, August 2005.

[24] M. Zinn, O. Khatib, B. Roth, J.K. Salisbury, "Playing it safe [human-friendly robots]", *Robotics & Automation Magazine*, IEEE Volume 11, Issue 2, June 2004 Page(s):12 - 21

Close Encounters: Spatial Distances between People and a Robot of Mechanistic Appearance [*]

Michael L Walters, Kerstin Dautenhahn, Kheng Lee Koay, Christina Kaouri, René te Boekhorst,
Chrystopher Nehaniv, Iain Werry, David Lee

{M.L.Walters, K.Dautenhahn}@herts.ac.uk
Adaptive Systems Research Group, School of Computer Science
University of Hertfordshire
College Lane, Hatfield, Herts, AL10 9AB, UK

Abstract **- This paper presents the results from two empirical exploratory studies of human-robot interaction in the context of an initial encounter with a robot of mechanistic appearance. The first study was carried out with groups of children, and the second with single adults. The analysis concentrates on the personal space zones and initial distances between robot and humans, the context of the encounters and the human's perception of the robot as a social being. We discuss the results of these observations and analyses, and also compare the child and adult data. The child groups showed a dominant response to prefer the 'social zone' distance, comparable to distances people adopt when talking to other humans. From the single adult studies a small majority preferred the 'personal zone', reserved for talking to friends. However, significant minorities deviate from this pattern. Implications for future work are discussed.**

Index Terms - Robot, Human-Robot Interaction, Social Spaces, Distances, Social Robot.

I. INTRODUCTION

The field of research into social and personal spaces with regard to robots, designed for use in the home, is a particular area of research within the wider field of Human - Robot Interaction (HRI). Although speech is an incidental part of these interactions, the main emphasis of this research is on the physical, spatial, visual and audible non-verbal social aspects of robots interacting socially with humans. An excellent overview of socially interactive robots (robots designed to interact with humans in a social way) is provided in Fong et al. [1]. As the study of socially interactive robots is relatively new, experimenters in the field often use existing research into human-human social interactions as a starting point. Hall [2] provided the original basis for research into social and personal spaces between humans, and later work in psychology has demonstrated that social spaces substantially reflect and influence social relationships and attitudes of people. Embodied non-verbal interactions, such as approach, touch, and avoidance behaviors, are fundamental to regulating human-human social interactions [3], and this has provided a guide for more recent research into human reactions to robots [4-7]. While the methods used to study human-human interaction may be relevant to this type of study, and the aim of many robot designers is to create robots that will interact socially with humans, it is probable that humans will not react socially to robots in *exactly* the same way that they react to other humans [8-12]. Previous work has generally assumed that robots are perceived as social beings and that humans will respond to a robot in a similar way, for example, as to a pet, another human, or even as to a child or infant. Evidence exists that humans do respond to certain social characteristics, features or behaviors exhibited by robots [13-15].

The *research hypothesis* advanced for empirically testing human-robot social space zones was that human-robot interpersonal distances would be comparable to those found for human-human interpersonal distances [2, 3]. The generally recognized personal space zones between humans are well known and are summarized (for northern Europeans) in Table 1 [16], which summarizes Hall's original distances.

TABLE 1
HUMAN-HUMAN PERSONAL SPACE ZONES

Personal Space Zone	**Range**	**Situation**
Close Intimate	0 to 0.15m	Lover or close friend touching
Intimate Zone	0.15m to 0.45m	Lover or close friend only
Personal Zone	0.45m to 1.2m	Conversation between friends
Social Zone	1.2m to 3.6m	Conversation to non-friends
Public Zone	3.6m +	Public speech making

We expect that in scenarios designed for direct human-robot interaction, people would assume distances that correspond to the Social or Personal zones (similar to the distances people use when having face-to-face conversations), thus treating the robot as a *social being*.

II. THE STUDIES

Two exploratory studies were carried out using commercially available, human-scaled, PeopleBot^{TM} robots. The first study took advantage of a larger software evaluation event, run by the FP5 European Project VICTEC [17], by providing 30-minute sessions for 24 groups of 10 children

---

[*] The work described in this paper was conducted within the EU Integrated Project COGNIRON ("The Cognitive Robot Companion") and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020.

involving interactive games with a PeopleBot^{TM} robot. The second study involved single human subjects interacting with the PeopleBot^{TM} robot in simulated living room scenarios. Prior to both studies, initial social space and comfort distance observations and measurements were carried out providing the main focus of this paper.

III. THE CHILD STUDY

A. *Experimental Setup and Procedure*

The robots used were commercially available PeopleBot robots fitted with a lifting arm, a pink hand, and a small basket which was used to hold small presents (Fig. 1). The arm could be raised or lowered under program control. The experiment was performed in an enclosed area of 6m x 6m which was marked out from the centre with a series of concentric circles at 0.5m radii intervals.

Fig. 1: The PeopleBot^{TM} robot fitted with arm and hand. This robot was used in the children's study.

The robot was positioned initially at the center of the circles, so an observer was able to use these to estimate the initial distance and relative orientation of members of each group; either directly or from the video recording of the session. The robots were controlled in a semi-autonomous manner, with the Wizard of Oz (WoZ) [18] operators hidden in an adjoining third room along with necessary equipment. The sessions were coordinated by an experimenter and followed the same overall format outlined here:

1) The children entered the room and each child was given a numbered sticker that was attached to their clothing so that the children could be tracked through the experiment.

2) An initial opinion questionnaire was administered before the children saw the robot and also asked for their genders and tracking numbers.

3) The robot was then uncovered and the experimenter let the children move around the robot without giving them any indication of where they should position themselves. Once settled, usually after a period of approximately 1 minute, each child's relative position and orientation towards the robot was recorded on paper record charts (cf. Fig. 2). The initial distances were estimated to the nearest 0.5m circle marking on the floor, giving an accuracy of ±0.25m. Also recorded was whether a teacher was present at the session, along with any other relevant observations.

4) Two interactive games were then played before a final questionnaire was administered. (These latter parts of the session were separate experiments and are not here).

The position and orientation measurements were later checked and verified against the video record after the session. This initial position information was recorded before the children had participated in the games (so before any interaction took place) and before the children had actually seen the robot move. The robot was stationary, though powered up and activated. Therefore, noises from the sonar range sensors and motors were audible throughout the game area.

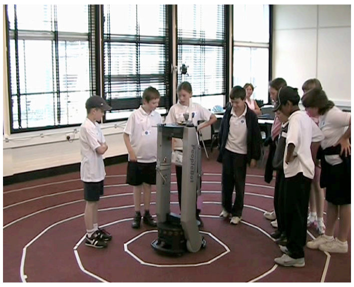


Fig. 2: A group of children take up their positions relative to the robot on their first encounter.

The PeopleBot^{TM} robot was mechanistic in appearance, so the only visual cues that indicated the front of the robot were:

1) The direction which the robot moved, either forwards or reverse, gave an indication of possible front and rear ends of the robot. This was not apparent until the robot moved, which did not occur at this stage of the test. Therefore, it would not be a factor to consider in this part of the study.

2) The Camera was mounted on top and to the front edge of the robot and pointed forward when the robot was activated but was stationary.

3) The PeopleBots used in the experiment were fitted with a simple arm, on the right hand side. It was in its lowered position at this stage of the experiment. On the left hand side, the robots were fitted with a basket (empty at this stage) to hold presents which would be given during the course of the later game experiment.

B. *Results from the Child study*

From the total sample of 196 children, only 131 (71 boys and 60 girls) have been included in the analysis. This was either because some of the children had been told explicitly where to stand initially by a teacher or adult, or were not given an opportunity to take up their initial positions. The initial distance results are summarized below and in Fig. 3.

Initial Distances; All - Mean = 1.73m, Median = 1.75m, St. Dev. = 0.73m

Girls Initial Distances - Mean = 1.74m, Median = 1.75m, St. Dev. = 0.61m

Boys Initial Distances - Mean = 1.72m, Median = 1.25m, St. Dev. = 0.73m

The children initially tended to place themselves at an overall mean distance of 1.75m (St Dev = 0.73) which is consistent with the social distance which would be used by humans to communicate with non-friends, and ranges from 1.2m to 3.6m (social zone, cf. Table 1).

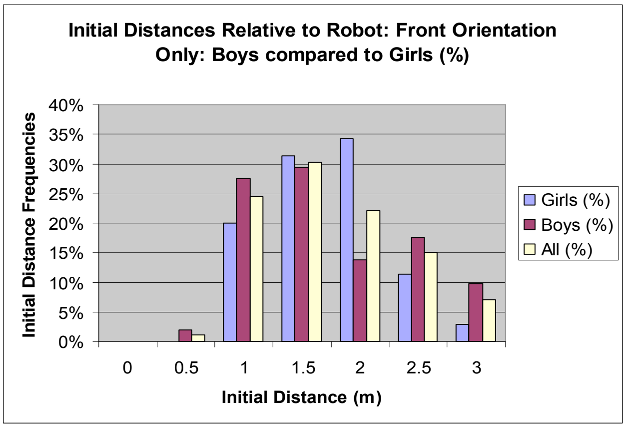


Fig 3: Frequencies of initial distances relative to robot: Front orientation only, for boys, girls and all children

The implication is that most of the children might have related to the robot as a social entity, even though the PeopleBots used for the study only had one arm (which was the only explicit anthropomorphic feature). The robot's onboard video camera may also have acted as a focus of the children's attention, the children may have been eager to interact due to the play context or because of the detailed preparation necessary for the event to take place (a school excursion), they may have been primed to expect an interaction with the robot to take place.

The initial orientation of each child was estimated by which quarter each child was initially positioned in relative to the robot, and recorded as front, right, left or back. The results are summarized in Table 2. The initial distance and orientation results presented here suggest that there is a strong tendency for a majority of just over half of the children (53%) to position themselves to the front of the robot initially. There was also an indication that proportionally more boys than girls (59% to 47%) positioned themselves at the front of the robot. However, although the mean distance of the boys from the robot (1.72m, St Dev = 0.73) is similar to the girls mean distance (1.74m, St Dev = 0.61), it can be seem from the standard deviation values and from the chart (Fig. 3) that the boys actually tended to place themselves initially either relatively closer to, or further away from, the robot than the girls who tended to exhibit a more compact normal distance frequency distribution.

From the number of children who positioned themselves at the front of the robot (53%), one might infer that the camera or the pointer (or both together) are powerful attractors of the children's initial attention, even though the camera, arm and robot were stationary. There may also be a weaker indication that the stationary arm pointer possibly had some effect in causing some children to prefer positions on the robot's right side (21%) as opposed to left side (12%) or behind the robot (12%). However, the entrance to the game area was to the right of the robot so this may possibly have affected this observed right-left preference. Further experiments should control for this and also for the initial orientation of the robot.

TABLE 2
CHILDREN'S INITIAL ORIENTATION MEASUREMENTS

Children's Initial Orientations Relative to PeopleBot Robot.	
Children in front of robot	70 (53%)
Children to right of robot	27 (21%)
Children to left of robot	16 (12%)
Children behind robot	16 (12%)

IV. THE ADULT STUDY

A. *Experimental Setup and Procedure*

This study was an exploratory investigation and involved twenty-eight single subject sessions with individual adults interacting with a single robot in simulated living room scenarios. These experiments applied a human-centered perspective; which is concerned with how people react to and interpret a robot's appearance and/or behavior, regardless of the cognitive processes that might happen inside the robot (robot-centered perspective). A large conference room was converted and furnished to provide as homely a environment as possible. Adjacent was an enclosed section where the WoZ robot operators and equipment were housed.

The subject sample set consisted of 28 adult volunteers [male: N: 14 (50%) and female: N: 14 (50%)] recruited from the University. A small proportion (7%) was under 25 years of age, but no one younger than 18 took part. Approximately 43% were 26-35 years old, 29% 36-45 years old, 11% 46-55 years old and 11% were over 56 years of age. 39% of the participants were students, 43% academic or faculty staff (e.g. lecturers, professors) and 18% were researchers in an academic institution. Approximately 50% came from a robotics or technology-related department (e.g. computer science, electronics and engineering), and 50% came from a non-technology related department, such as psychology, law or business. All subjects completed consent forms and were not paid for participation, but at the end of the trial they were given a book as a present.

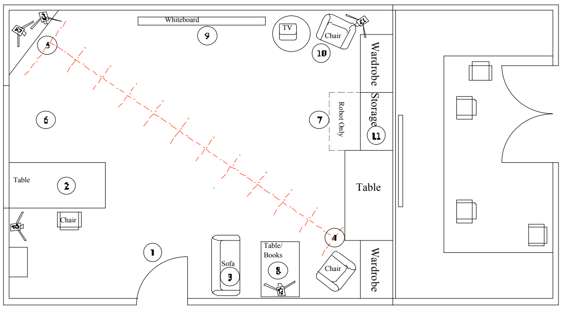


Fig. 4: Plan view of simulated living room layout. Comfort distance tests carried were out along the marked diagonal line.

The initial distance measurements were performed before a separate experimental session involving human-robot interactions in task based scenarios. Scale marks were made at 0.5m intervals along the diagonal of the room (Figs. 4 and 5) and the human-robot comfort and approach distances were estimated from the video records, rather than making intrusive measurements or notes during the sessions. The robot's arm was adapted so that it could pick up and carry small palettes which contained items to be brought to the human subject later on in the task scenarios (Fig. 6) (Note; The hand was not as anthropomorphic in appearance as that used for the child study). Each experiment session followed the same format:

1) Entry to room and introduction of robot

2) Co-habituation and initial questionnaires. While the subject was completing the questionnaires, the robot wandered randomly around the test area. Unlike the first study the subject was allowed to acclimatize to the robot for five to ten minutes prior to the distance tests.

3) Comfort and social distance tests.

4) Various other HRI task scenarios and questionnaires. (These latter parts were carried out for separate HRI investigations and are therefore not considered in this paper).

Fig. 5: Views of simulated living room showing robot and the 0.5m scale marked diagonally on the floor

The experiments were supervised by an experimenter who introduced and explained the tests to be carried out to the subject. Otherwise, she interfered as little as possible with the actual experiment. For measuring the human subject's comfortable distance when approaching the robot, the robot was driven to point 5 (next to the corner table) and turned to face along the distance scale towards point 4 (Fig. 4). The subject was told to start at point 4 and to move towards the robot until he or she felt that they were at a comfortable distance away from the robot. Next, they were told to move as close to the robot as they physically could, then to move away again to a comfortable distance. They were then told to repeat these steps once again as a consistency check. The comfortable approach, closest physical and comfortable withdrawal distances were measured for each of the two tests to the nearest 0.25m (accuracy ±0.125m) by later close observation of the video records.The next part of the comfort distance tests was to measure the subject's comfort distance with the robot moving from point 5 towards the subject. The subject was told to stand at point 4, and the robot moved directly towards him or her. The subject was told to say, "Stop", when the robot was as close as the subject desired. The distance of the robot when the subject said, "stop" was estimated later, and recorded, from close observation of the video records.

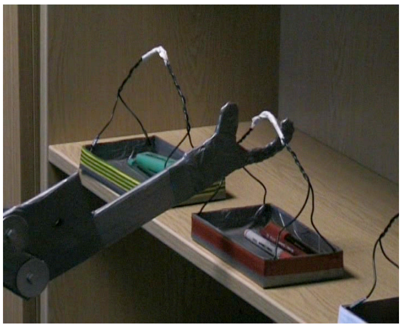


Fig 6: Detail showing the robot's arm and hand used in the study with adult subjects.

B. *Results from the Adult Study*

The means of the four robot comfortable approach distance results obtained was calculated for each subject and the frequency histogram was plotted, with the ranges set at 0.25m intervals. The results are presented in the chart in fig. 7. Approximately 40% of subjects approached the robot to a distance of 0.5m or less. When the robot approached a human, the anti-collision safety system prevented it moving closer than 0.5m. Due to safety concerns this system must be kept operational.

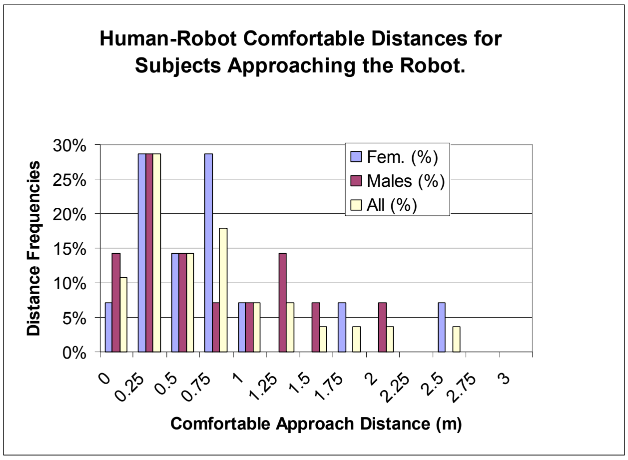


Fig. 7: Comfortable approach distance frequencies for subjects approaching the robot.

It can be seen that approximately 40% of the subjects also allowed the robot to approach right up to this 0.5m limit. That they did not stop the robot from physically approaching so closely to them indicates that the robot did not make them feel threatened or uncomfortable. When asked later if they felt uncomfortable while standing in front of the robot most subjects (82%) indicated that they were not uncomfortable. Also, as less than 20% indicated that they wanted a robot for a friend or companion, these close approach distances did not express the subjects' wish to be intimate with the robot. That many of the subjects approached the robot closely, and tolerated a relatively close approach implies that they might not see the robot as a social entity in the same way that they would perceive another human. If another, unfamiliar human (a stranger) was to approach to the same close distances; most humans would feel distinctly uncomfortable and threatened. Interestingly, there were a small number of subjects (approximately 10%) who were uncomfortable in letting the robot approach closer than the far end of the social zone (>1.2m and <3.6m), which is usually reserved for conversations between humans who are strangers to each other.

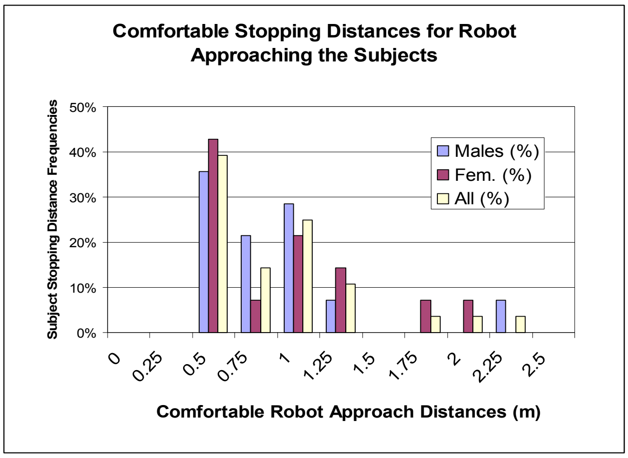


Fig. 8: Comfortable approach distance frequencies for the robot approaching the Subjects.

V. COMPARISON AND DISCUSSION OF RESULTS FROM BOTH STUDIES

While the experimental setups of both studies were very different, comparisons between the comfort zones measured in both studies can highlight results which generalize across the two different experimental setups. In both studies, for the children and adult subjects, a majority took up an initial position relative to the robot which was consistent with treating the robot as a social being with respect to accepted human-human social space zones. In both studies, the floor was marked with scale marks in order to aid the distance measurements and this may have influenced the distance results obtained. A major difference between the two social distance studies was that the children were interacting with the robot in groups, whereas the adults were interacting individually with the robot. It is very likely that the children took cues from, and were interacting with each other as well as the robot; cf. a discussion of social facilitation effects in Woods et al. [19]. Almost all the children generally took up a mean distance which would, amongst humans, be reserved for talking or interacting with strangers or other non-friends. However, most adults took up a distance which in a human-human context would be used for talking with friends. Generally, these results support our initial research hypothesis, namely that distances used in direct human-human social interaction can apply to robots. In both cases, this could however simply be a convenient distance for viewing the robot, so more tests are required to confirm the reasons for these observations. A small proportion of each group took up an initial distance as far from the robot as the limited space allowed.

Interestingly, amongst the adults, there was a sizable minority (approximately 40%) who took up an initial position relative to the robot which was so close that it would be classified as that reserved for intimate lovers or friends. This probably means that those subjects did not see or treat the robot as a social being. Pamela Hinds and colleagues [20] have studied the effect of robot appearance on humans carrying out a joint task with a robot. They found that humans treat mechanistic looking robots in a subservient way (i.e. less socially interactive) compared to more humanoid looking robots. Also expectations are lower as regards abilities and reliability for mechanistic looking robots.

The PeopleBot^{TM} robots used in the studies were fitted with a moving articulated arm. However, they are still very mechanistic in their appearance, so it is probable that many subjects in the adult experiment simply did not recognize the robot as anything more socially interesting than any other household object or machine (such as a refrigerator or television). Amongst the children, only a very few went so close to the robot initially. Results indicate that they possibly saw the robot as a social entity. This may reflect their different expectations, lesser discrimination and self-consciousness in interacting with the robot in a play context.

For both child and adult studies the social distance experiments were performed before any other interactions had taken place. With more opportunity for habituation, the perception of the subjects may have changed over the course of the experiments. It would be useful to perform distance experiments both before and after exposure to robot scenarios to see how subjects' perceptions change with both short and longer term exposure to robots. There is a need, therefore, to perform long-term studies (over periods of longer than one hour) and repeated exposure of the subjects to the robot over longer periods of time.

The adult study did not consider the initial orientation of the subjects to the robot, due to lack of space in the experimental room, but the children's study did gain results that indicated that the only two possible anthropomorphic features which distinguished the front and back of the robot, the hand/arm and the camera, did probably exert an effect on where the children chose to orientate themselves when initially encountering the robot. There are also some indications that the arm and hand may also exert a right hand bias to the children's initial orientations, though this needs further study to confirm as the entry to the game arena was also to the right of where the robot was positioned initially.

VI. CONCLUSIONS

We cannot claim that the results gained as part of these two studies, using PeopleBots, can be generalized to any other type of robot or to any other context/scenario. The PeopleBots are mechanistic in appearance. These results could only possibly be extrapolated to include similar other robots.

Interestingly, substantial individual differences have been found in how people behaved towards the PeopleBot robot, although most of the human subjects participating in the studies seemed to be receptive to treating the robot as a (limited) social entity after only a short period. It seems that children in particular are more overall more accepting and approving of robots than the adult subjects studied. The social distance results so far have indicated that a substantial minority of adults (40% in our adult sample) do not seem to perceive the PeopleBot^{TM} robot as a social being at first encounter. There was a small but significant proportion of both children and adults, in the two studies documented here, who seem to be uncomfortable in the presence of the PeopleBot^{TM} robot.

There is a need for long term trials with a variety of types of robots in order to determine which social features are most effective at making human robot interaction robot more efficient and useful to humans. The CERO robot assistant study [21], the Robovie peer tutor robot trials with children [22], and trials involving children with autism interacting with a humanoid robot [23] are examples of the few published works which describe studies involving long term periods of humans interacting with robots. Different robot social models, perhaps with very different initial personalities, may be more acceptable to different users (e.g. a discreet servant or even a silent servant, with no obvious initiative or autonomy). Our results suggest that it probably cannot be assumed that people automatically treat robots socially, apart from simple elements of anthropomorphism cf. Reeves and Nass [24]. A user friendly robot should automatically refine and adapt its social model (personality) over a longer period of time, depending on information about and feedback from users and the robots own autonomous learning system. For example, adjustments of social distances according to a user's personality traits (as proposed in [25]) is a promising direction towards a true robot companion that needs to be individualized, personalized and adapt itself to the user [26].

ACKNOWLEDGEMENTS

Thanks go to Ben Robins, Alexander Klyubin, Lars Olssen and Aris Alissandrakis for their help in carrying out the trials.

REFERENCES

[1] T. Fong, I. Nourbakhsh, and K. Dautenhahn, (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, vol. 42, pp. 143-166, 2003.

[2] E.T. Hall, (1966). *The Hidden Dimension: Man's Use of Space in Public and Private*. The Bodley Head Ltd, London, UK. 1966

[3] E. T. Hall, (1968). Proxemics. *Current Anthropology* 9(2-3): 83-108. 1968

[4] J. Goetz and S. Kiesler, (2002). Cooperation with a Robotic Assistant *Proc. CHI'02 Conference on Human Factors in Computing Systems*, New York, USA, 2002.

[5] A. Paiva, J. Dias, D. Sobral. R Aylett, P Sobreperez, S Woods et al, (2004). Caring for Agents and Agents that Care: Building Empathic Relations with Synthetic Agents. *Third International Joint Conference on Autonomous Agents and Multiagent Systems* (AAMAS'04), pp. 194-201. July 19 - 23, 2004

[6] M. Scopelliti, M. V. Giuliani, A. M. D'Amico, and F. Fornara, (2004). "If I had a robot at home.... Peoples' representation of domestic robot". in *Designing a more inclusive world*, S. Keates, J. Clarkson, P. Langdon, and P. Robinson, Eds. Cambridge, UK: Springer, 2004, pp. 257-266.

[7] S. Woods, K. Dautenhahn, J. Schulz, (2004). The Design Space of Robots: Investigating Children's Views. *Proc. IEEE RO-MAN Conference, 13th IEEE Int. Workshop on Robot and Human Interaction Communication*. Kurashiki, Japan. Sept 2004.

[8] D. A. Norman, (1994). How might people interact with agents, *Communications of the ACM*, vol. 37, pp. 68-71, 1994.

[9] D. C. Dryer, (1999). Getting personal with computers: How to design personalities for agents*., Applied Artificial Intelligence*, vol. 13, pp. 273-295, 1999.

[10]Z. Khan, (1998). Attitude towards intelligent service robots. NADA KTH, Stockholm 1998

[11]K. Dautenhahn, (2002): Design Spaces and Niche Spaces of Believable Social Robots. *Proc. 2002 IEEE Int. Workshop on Robot and Human Interactive Communication ROMAN 2002*, 25-27 September, Berlin, Germany, pp. 192-197, IEEE Press.

[12]K. Dautenhahn, B. Ogden, T. Quick (2002). From Embodied to Socially Embedded Agents - Implications for Interaction-Aware Robots, *Cognitive Systems Research* 3(3), pp. 397-428)

[13]C. L. Breazeal, (2002). *Designing sociable robots*. Massachusetts: The MIT Press, 2002.

[14]B. Friedman, P. H. Kahn (Jr.), and J. Hagman, (2003). "Hardware companions? What online AIBO discussion forums reveal about the human-robotic relationship," *Digital Sociability*, vol. 5, pp. 273-280, 2003

[15]Sony, (2004). "Sony Entertainment Robot Europe http://www.aibo-europe.com/".

[16]D Lambert, (2004). *Body Language*. HarperCollins.

[17]VICTEC, (2003). VICTEC Project website available at http://www.victec.org, 2003.

[18]A. Green, H. Hüttenrauch, K. Severinson Eklundh (2004). Applying the Wizard of Oz Framework to Cooperative Service discovery and Configuration. *Proc. IEEE Ro-man 2004*, *13th IEEE International Workshop on Robot and Human Interactive Communication*, Okayama, Japan, IEEE Press.

[19]S Woods, K Dautenhahn, C Kaouri, (2005) Is Someone Watching Me? – Consideration of Social Facilitation Effects in Human-Robot Interaction Experiments. *Proc. IEEE CIRA 2005*.

[20]P Hinds, T Roberts, H Jones, (2004). Whose Job Is It Anyway? A Study of Human-Robot Interaction in a Collaborative Task. *Human-Computer Interaction, 2004* Vol. 19 pp151-181.

[21]K Severinson-Eklundh, Anders Green, Helge Hüttenrauch, (2003). Social and Collaborative Aspects of Interaction with a Service Robot. *Robotics and Autonomous Systems* 42 (2003) pp. 223-234.

[22]T Kanda, T Hirano, D Eaton, (2004). "Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial". *Human-Computer Interaction, 2004*, Vol. 19, pp. 61-84.B

[23]Robins, K Dautenhahn, R te Boekhorst, A Billard, (2004) "Effects of repeated exposure to a humanoid robot on children with autism". In S. Keates, J. Clarkson, P. Langdon and P. Robinson (Eds.) *Designing a More Inclusive World*, Springer Verlag, London, pp. 225-236.

[24]B Reeves, C Nass, (1996). *The Media Equation: How people treat computers, television and new media like real people and places.* Cambridge University Press 1998. ISBN 157586052x

[25]M L Walters, K Dautenhahn, R te Boekhorst, K L Koay, C Kaouri, S Woods, C L Nehaniv, D Lee, I Werry, (2005) The Influence of subjects' Personality Traits on Personal Spatial Zones in a Human-Robot Interaction Experiment. *Proc. IEEE Ro-man 2005, 14th Annual Workshop on Robot and Human Interactive Communication,* 13-15 Aug 2005. Nashville, Tennessee, USA.. IEEE Press, pp. 347-352.

[26]K Dautenhahn, (2004). Robots We like to Live With?! – A Developmental Perspective on a Personalized Life-Long Robot Companion. *Proc. IEEE Ro-Man 2004, 13th Annual Workshop on Robot and Human Interactive Communication,* September 20-22, Kurashiki, Okayama, Japan. IEEE Press, pp. 17-22.

Practical and Methodological Challenges in Designing and Conducting Human-Robot Interaction Studies

Michael L. Walters, Sarah Woods, Kheng Lee Koay, Kerstin Dautenhahn
{m.l.walters, S.N.Woods, K.L.Koay, K.Dautenhahn}@herts.ac.uk
Adaptive Systems Research Group
School of Computer Science
University of Hertfordshire
College Lane, Hatfield, AL10 9AB

Abstract

Human-robot interaction is a rapidly growing research area which more and more roboticists and computer scientists are moving into. Publications on work resulting from such studies rarely consider in detail the practical and methodological problems encountered. This paper aims to highlight and critically discuss such problems involved in conducting human-robot interaction studies. We provide some examples by discussing our experiences of running two trials that involved humans and robots physically interacting in a common space. Our discussion emphasises the need to take safety requirements into account, and minimise the risk of physical harm to human subjects. Ethical considerations are considered, which are often within a formal or legal framework depending on the host country or institution. We also discuss future improvements for features of our trials and make suggestions as to how to overcome the challenges we encountered. We hope that the lessons learnt will be used to improve future human-robot interaction trials.

1 Introduction

In the course of our research for the COGNIRON Project [2005], we are primarily interested in the research area of Human-Robot Interaction (HRI), in particular with regard to socially interactive robots. An excellent overview of socially interactive robots (robots designed to interact with humans in a social way) is provided in Fong et al. [2003]. As we are primarily studying the human perspective of human-robot interaction, human scaled robots in live trials within a human orientated environment were required.

Fig. 1: Children playing games with the University of Hertfordshire PeopleBot^{TM} at the London Science Museum event in October 2004.

Other researchers that have conducted similar human centred trials with human sized robots include Dario et al. [2001], Severinson-Eklundh et al. [2003], Kanda et al. [2004] and Hinds et al. [2004]. To date, we have conducted two human-robot trials with human scaled PeopleBot^{TM} robots. One trial involved a single robot interacting with groups of children in a game scenario. The trials took advantage of a software evaluation event at the University of Hertfordshire, hosted by the Virtual ICT Empathic Characters (VICTEC) project [VICTEC, 2003]. The other trial involved individual adults interacting with a robot in various contexts and situations, within a simulated domestic (living-room) environment. We have also participated in other displays and demonstrations which have involved robots interacting in the same physical space as one or more humans. In particular, we successfully ran interactive games for groups of up to 40 children at a time, at a major public event at the Science Museum in London [BBC Science News, 2004]. The PeopleBot^{TM} robots have also been demonstrated on several occasions during open days at the University of Hertfordshire. This paper will present some of the methods we have developed and critically discuss the various trials and events we have been involved with to date.

2 Planning, Legal and Safety

Before running a trial involving humans and robots physically interacting, certain legal and ethical issues must be satisfied. At this stage it is good practice, and in the UK a legal requirement under the Management of Health and Safety at Work Regulations 1999, to carry out a risk assessment for all work activities involving employees or members of the public [Crown Copyright, 2003]. These first activities are considered here.

2.1 Legal and Ethics Approval

Many institutions, including the University of Hertfordshire [UPR AS/A/2, 2004], require that an Ethics Committee must give approval for all experiments and trials involving human subjects. Usually, this approval is gained by submitting a (written) description of the trials or experiments to be performed to the committee. The Ethics committee will then consider the proposal, and may modify, request further clarification, ask for a substantial rewrite, or even reject the proposal outright on ethical grounds. In general, the Ethics committee will make possible objections on the following grounds:

Privacy – If video, photographic or records of personal details of the subjects are being made and kept, the committee will be concerned that proper informed consent is given by subjects, any personal records are securely stored and will not be misused in any way. If personal data is to be held on a computer database, then the legal requirements of the Privacy and Electronic Communications Regulations [Crown Copyright, 2003] must be adhered to. If any public use of the video or photographs is to be made for conferences or publicity purposes, then participants must give explicit permission.

Protection of minors and vulnerable adults – In the UK it is a legal requirement (Protection of Children and Vulnerable Adults Order, 2003) that anyone who works with children or vulnerable adults must have their criminal record checked. In the UK, anyone under 18 is classed as a child in this context, and the term vulnerable adult includes the infirm or elderly in a care situation. Regulations in many other countries in Europe are less strict, but if experiments or trials are planned to involve children or vulnerable adults, then any legal implications or requirements must be considered. For example, not gaining the appropriate checking of criminal records could lead to a situation where subjects who are keen to participate in a study need to be turned down. Given the general problem in recruiting a sufficiently large sample of human subjects, this could potentially cause problems.

Mental or emotional stress and humiliation – The trials should not give rise to undue mental or emotional stress, with possible long-term repercussions. Where an experimental situation is actually designed to put a subject under stress intentionally, it may not be possible to avoid stressing the subject. The Ethics Committee will want to be satisfied that if any mental or emotional stress is suffered by subjects, it is justified and that no after effects will be suffered by subjects. In our own studies we were interested in how subjects 'spontaneously', or 'naturally' behaved towards robots, so we had to carefully design the scenarios in order to be on the one hand controlled enough to be scientifically valuable, but on the other hand open enough to allow for relaxed human-robot interactions. It is advised to include a statement in the consent form which points out that the subject can interrupt and leave at any stage during the trial for whatever reasons, if he or she wishes to.

Physical harm – Practically all experiments that involve humans moving will involve some degree of risk. Therefore, any human-robot trials or experiments will pose some physical risk for the subjects. The Ethics Committee will want to be satisfied that the proposal has considered any potential physical risks involved. The subjects' safety is covered in more detail below.

2.2 Safety

Robot-Human Collision Risk - For trials involving humans and robots, the obvious immediate risk is the robot colliding with a human subject, or vice versa. The robots that we used in our trials are specifically marketed for the purpose of human-robot interaction studies. In order to alleviate the risk of the robot colliding with human subjects, two strategies were adopted:

Overriding anti-collision behaviour – The PeopleBot^{TM} robots we use can have several behaviours running at the same time as any top-level program. This is a natural consequence of the PeopleBot^{TM} operating system which follows the principles of the subsumption architecture expounded originally by Rodney Brooks [1991]. Many other commercially available robot systems have similar programming facilities. We always had basic collision avoidance behaviours running at a higher priority than any task level program. This means that no matter what the task level program commands the robot to do, if a collision with an object is imminent, the underlying anti-collision behaviour cuts in. Depending on the form of the hazard and the particular safety behaviour implemented, the robot will either stop or turn away from the collision hazard. The lower priority task level programs include both those that provide for direct or semi-autonomous remote control by Wizard of Oz (WOZ) operators [Maulsby et al. 1998] and also fully autonomous programs. We have found that the sonar sensors used by the PeopleBot^{TM} are very sensitive to the presence of humans. However, some common household objects, especially low coffee tables, are not so readily picked up by the sensors. By judicious placing of objects that are readily sensed, such as boxes, footballs, cushions etc, it is possible to create a trial environment where it is literally impossible for the robot to collide with any object. For example, we adopted this strategy to avoid the robot bumping into the table where the person was sitting (see fig.2).

Fig. 2: A subject sitting at the desk, showing a box placed under the table to create a target for the robot's sonar sensors.

Monitoring by the WOZ operators – Even while the robot is running a fully autonomous program, a WOZ operator (see section 3.1) monitors discreetly what is happening. The robot's underlying safety behaviours include the overriding ability for the WOZ operator to stop the robot immediately by remote wireless link if it is perceived that the robot poses a risk to a human at any time. There is also a large red emergency stop button on the robot, which is hardwired, providing an independent failsafe method to stop the robot. Simply pressing the button cuts the power to all the robot's motors. This is simple enough for non experts to operate, and will work even if the robots control software crashes or fails to respond. Anyone who is physically close enough (i.e. in perceived danger) to the robot can access the button.

In our trials, only during the software development process of the program has it been necessary for WOZ operator or others to initiate a stop; mainly to avoid the robot damaging itself rather than actually posing a threat to humans in the vicinity. During our human-robot interaction trials, the underlying safety behaviour has proved to be both robust and reliable in detecting and avoiding collisions with both children and adults. The actual robot programs have been heavily tested in the physical situations for all the trials we have run. This is necessary as knowing how the robot will respond in all physical circumstances is critical for the safety of the participants in any trial.

For the risk case of a human colliding with the robot, there is little action that can be taken by the robot to avoid a human. The robot moves and reacts relatively slowly, compared to the speeds achievable by a human. Therefore, it is up to the human to avoid colliding with the robot. Luckily, most humans are experts at avoiding collisions and we have found that none of our subjects has actually collided with the robot. In some of the trials with children it has been necessary to advise the children to be gentle or to move more carefully or slowly when near to the robot. We found that children will mostly take notice if the robot actually issues these warnings using the robot's own speech synthesis system.

Other Possible Risks to Participants - Our robot was fitted with a lifting arm, which had a small probability of causing injury to humans. The arm itself was made of coloured cardboard made to look solid, so it looked more dangerous than it actually was. Our main concern about the arm was if the 'finger' was accidentally pointed into a human's face or eyes. This risk was minimised by keeping the arm well below face level even when lifted. Other possible risks to participants that must be considered are those that would be present in any domestic, work or experimental situation. These include things such as irregular or loose floor coverings, trailing cables, objects with sharp or protruding edges and corners, risk of tripping or slipping, etc.

In our trial involving children, small prizes were given during and at the end of each session. We were advised against providing food (i.e., sweets) as prizes, as some children may have had allergies or diabetes which could be aggravated by unplanned food intake. We also never left subjects alone with a robot without monitoring the situation.

3 Experimental Implementation

When running a human-robot interaction trial, the question that must be addressed is how to implement the proposed robot functions and behaviour. There are two main methods for developing suitable robot features, functions and behaviour for trials where we are primarily interested in the human-centred perspective towards the robot or its function.

3.1 Wizard of Oz Methods

It is usually relatively quick to create a scenario and run the robot under direct WOZ operator control. This is a technique that is widely used in HRI studies as it provides a very flexible way to implement complex robot behaviour within a quick time-scale (Robins et al. 2004 and Green et al. 2004). The main advantage is that it saves considerable time over programming a robot to carry out complex interactions fully autonomously. However, we have found that it is very tiring for the WOZ operators to control every aspect of the robot's behaviour, especially in multi-modal interactions and scenarios. It usually requires two operators, one for controlling movement and one for speech, in order to maintain reasonable response times during a trial. It is also difficult to maintain consistency between individual trial sessions. Practise effects are apparent as the operators become better at controlling the robot at the particular task scenario through the course of a series of trials. Practise effects can be minimised by thoroughly piloting the proposed scenario before carrying out 'live' trials.

Fig. 3: The Wizard of Oz operators and control room area for human-robot interaction trials at the University of Hertfordshire in 2004.

3.2 Autonomous Robot Control

The other robot control method is to pre-program the robot to run all functions autonomously. Obviously this method overcomes the problems of operator tiredness and consistency, but implementing complex autonomous behaviour is very time-consuming. However, if trials are testing complex human-robot social behaviours, or implementing desired future robot capabilities, it will not be technically feasible at present to program a robot to act fully autonomously. In accordance with the COGNIRON project aims, we are studying scenarios that go "beyond robotics". For this we have to project into the future in assuming a robot companion already exists that can serve as a useful assistant for a variety of tasks in people's homes. Realistically, such a robot does not yet exist.

The PeopleBot^{TM} robots have a sophisticated behaviour based programming API called ARIA [ActivMedia Robotics, 2005]. This provides facilities to develop task control programs, which can be integrated into the ARIA control system. The actual task control program can be assigned a priority, which is lower than the previously mentioned safety behaviours (see section 2.2). Therefore, fundamental safety and survival behaviour, such as collision avoidance, emergency stop etc. will always take precedence over the actual task commands.

In practice, we have found that a mixture of autonomous behaviours and functions, and direct WOZ control provides the most effective means of generating the desired robot's part of the HRI. The basic technique is to pre-program the robot's movements, behaviours or sequences of movements, as individual sequences, gestures or actions that can be initiated by the WOZ operator. In this way the WOZ operator is able to exercise judgement in initiating an appropriate action for a particular situation, but is not concerned with the minute details of carrying that action out. The operator then is able to monitor the action for potential hazard situations and either stop the robot or switch to a more appropriate behaviour. Because the robot is actually generating the individual movements and actions autonomously, better consistency is ensured. Also, the temporal behaviour of a robot under WOZ or autonomous control is likely to differ significantly, so whenever possible and safe, autonomous behaviour is advantageous over remote-controlled behaviour.

Robot program development & pilot studies- When developing robot programs, which will be used to implement a HRI trial scenario, it is important to allow enough time to thoroughly practise the programs and scenarios thoroughly before the actual trials take place. Pilot studies should be conducted with a variety of humans, as it is easy for the programmer or operator to make implicit or explicit assumptions about the way that humans will behave in response to a given trial situation. Of course, humans all exhibit unique behaviour and can do unexpected things which may cause the robot program to fail.

The first trials we ran involved interactive game sessions with groups of children. These required the children to play two short games with the robot, a *Rotation game* and a *Wander game*. The game programs ran mostly autonomously, except for starting the respective game programs, and also at the end of each round where a winning child was selected manually by remote control. When developing the interactive game programs for the Science Museum visit, the games ran totally autonomously for the whole of each game session. The Science museum game program was more complex than the previous child group games programs as sensor interpretation was involved. However, because the Science Museum robot game program was fully autonomous, the pre-testing phase had to be much longer. The extra time was needed to empirically find out optimum action and response timings and durations, sensor levels and cues, and refining the program so that it worked properly with all the human test subjects.

For the single adult HRI trials, there were time limitations on setting up and implementing the experiment. The robot behaviour was implemented almost entirely by direct WOZ control (with overriding safety behaviour active). There was also limited time available for practicing the scenarios, which were to be implemented for the study. The only autonomous behaviours used for this study were the wandering behaviour, used for acclimatising the subject to the robot's presence, and the arm lift height, which was used to set the arm to the correct height for picking up special pallets which contained items that would be fetched by the robot at various times during the trial (fig 4).

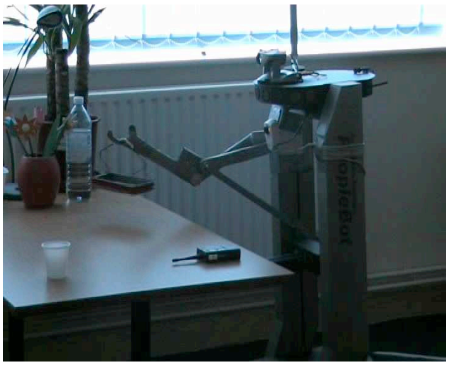


Fig. 4: The robot, fitted with a hook-like end-effector, was able to fetch small items in special pallets.

The WOZ operators were out of direct sight of robot and subject, and observed the scenarios via network video cameras placed around the room. The images from these were delayed by approximately 0.5 sec. There was also a direct, but restricted, video view from the robot camera which did not have any discernable delay. These factors made providing timely responses (comparable to human responses) to the subject very difficult for the WOZ operators. However, it can be argued that, in the near future at least, this is likely to be true of all robots, and this was a realistic simulation of likely future robot performance.

4 Video Recording

It is desirable to make a complete video record of the trials. Video footage is one of the primary means of gaining results for later analysis and validation of results. They can be used to validate data obtained by other means, e.g. from direct measurement, questionnaire responses, or recorded sensor data. Good video footage can provide time stamped data that can be used, processed and compared with future studies. However, in addition to the obvious advantages of video data, there are some drawbacks that researchers should be aware of at the outset of the design phase. Analysing video footage is an extremely time consuming process and requires thorough training in the application of the scoring procedures, which can be complex. Observations made from video footage are subjective and the observer may portray their own perceptions and attitudes into the date. For this reason, it is essential that a full reliability analysis of video data is carried out involving independent rating and coding by observers who were not involved with the study, and did not meet any of the participants.

4.1 Video camera types

We used two types of video cameras for recording our trials; tripod mounted DV camcorders, and network cameras. The DV camcorders record onto mini DV tape, which must then be downloaded onto a computer hard disk before further analysis can be performed. The network cameras have the advantage that they record directly to a computer's hard disk, so there is no tedious downloading later on. They do require some synchronising, converting and combining, but this can be done automatically in batches overnight. We have found that the DV Camcorders provide a better quality picture than the network cameras, with a synchronised soundtrack. While high quality video may not be strictly necessary for analysis purposes, it does allow high quality still pictures to frame-grabbed from the video recordings, which are invaluable for later writing up, papers and reports. It is also easy to create short videos to incorporate into presentations and demonstrations using standard video editing software.

It is advisable to use at least two camera systems for recording trials or experiments. If one camera fails, than there will another stream of video data available. It should be borne in mind that if a network camera fails, it may also lead to all the network cameras being bought down. Therefore, at least one camera should be a freestanding camcorder type, which stores the video data on (mini DV) tape.

Note, a similar backup strategy is also advisable as far as the robotic platforms are concerned. In our case, we had a second PeopleBot^{TM} in place, in the event that one robot broke down. Having only one robot available for the trials is very risky, since it could mean that a trial had to be abandoned if a robot fails. Re-recruiting subjects and properly preparing the experimental room is a very time-consuming activity, unless a permanent setup is available. This was not the case in our trials, where rooms were only temporarily available for a given and fixed duration (two weeks for the study involving children, 2 months for the adult study). Afterwards the setups had to be disassembled and the rooms had to be transformed back into seminar or conference rooms. This also meant that any phases of the trials could not be repeated. Therefore, it was essential to get it right first time despite the limited preparation time. This is a situation common to a University environment with central room allocations and usually few permanent large laboratory spaces suitable for studies with large human-sized robots.

4.2 Camera Placement

The placement of the cameras should be such that the whole trial area is covered by one or two views. For our first trial, we used two cameras placed in opposite corners of the room, both facing towards the centre of the room. As a result we recorded two views of the centre of the room, but missed out on what was happening at the edges of the room. A better way to position the cameras would have been to point the cameras to the right (or left) of room centre, with only a small view overlap in the centre of the room. This way, the two views also include the outer edges of the room. (See fig.5)

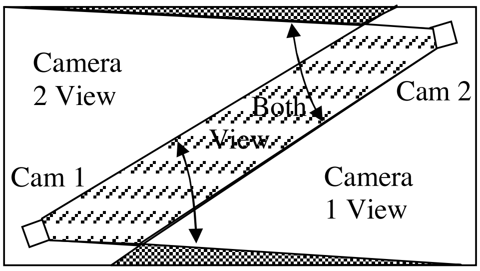


Fig. 5: Diagram showing correct placement of cameras to maximise coverage of room.

It is best to use two cameras to cover the entire area as shown in the diagram, with additional cameras to obtain detailed views of specific areas of interest. For example; when it is known that subjects will have to sit at a certain desk, which is in a fixed position, it is worth setting up an individual camera just to record that position in detail. Setting the correct height of the cameras is important to obtain a good view of the subjects face.

4.3 Distance Measurements

One main aspect of our trials was focused on examining the spatial distances between the robot and human subjects. Video images can be useful in estimating these distances. In both our child group and single adult trials, markings were made on the floor with masking tape to provide a method to estimate the position of the robot and the human subjects within the trial areas. However, these markings were visible to the subjects, and may possibly have influenced the positioning of the human subject during the course of parts of the experimental scenarios.

In the context of a study described by Green et al. [2004], a method was used that involved overlaying a grid of 0.5m squares onto still images of the floor of their trial area for individual frames from their video recordings. This method would allow the positions of the robot and subject to be estimated with a high degree of accuracy if it can be adapted for live or recorded video data. It would provide a semi-transparent grid metric overlaid onto the floor of the live or recorded video from the cameras. The possibility of visible floor markings affecting the positions taken by the subject would not happen. For future trials we will want to use such a 'virtual grid' on the floor of the recorded video data. We are currently evaluating suitable video editing software.

5 Subject's Comfort Level

For the adult trials, we experimented with a method of monitoring how comfortable the subject was while the trials were actually running. We developed a hand held comfort level monitoring device (developed by the first author) which consisted of a small box that could be easily held in one hand (see fig. 6). On one edge of the box was a slider control, which could be moved by using either a thumb or finger of the hand holding the device. The slider scale was marked with a happy face, to indicate the subject was comfortable with the robot's behaviour, and a sad face, to indicate discomfort with the robot's behaviour.

Fig. 6: Photograph of Hand Held Comfort Level Monitoring Device

The device used a 2.4GHz radio signal data link to send numbers representing the slider position to a

PC mounted receiver, which recorded the slider position approximately 10 times per second.. The data was time stamped and saved in a file for later synchronisation and analysis in conjunction with the video material. The data downloaded from the hand held subject comfort level device was saved and plotted on a series of charts. However, unexpectedly, the raw data was heavily corrupted by static from the network cameras used to make video recordings of the session. It has been possible to digitally clean up and recover a useful set of data. A sample of the raw data and the cleaned up version is shown in the figs. 7 and 8.

Many of the comfort level movements correspond to video sequences where the subject can be seen moving the slider on the comfort level device. This confirmed that the filtered files were producing a reliable indication of the comfort level perceived by the subject. For future trials, it is intended to incorporate error checking and data verification into the RF data transfer link to the recording PC in order to reduce problems with static.

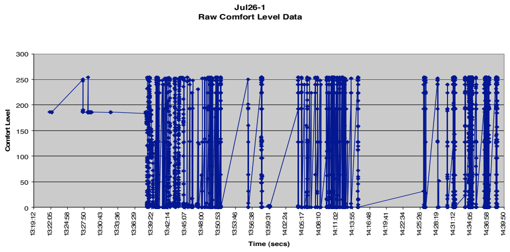


Fig. 5: Raw Data as Received from Handheld Comfort Level Monitoring Device

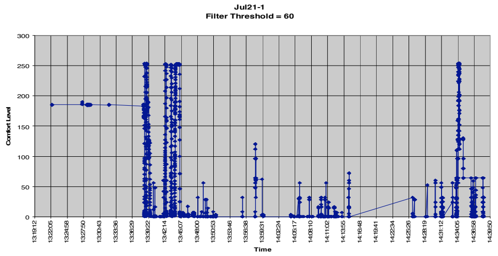


Fig. 6: Digitally Filtered Data from Handheld Comfort Level Monitoring Device

6 Questionnaires

For both interactive trials, subjects were asked to complete questionnaires. For the child-robot interactions only five minutes at the beginning, and five minutes at the end of each session were available. Due to limited time, only basic information was obtained, such as gender, age, approval of computers and robots, and how they liked the interactive session. For the adult study, the questionnaires were much more comprehensive. The time taken for the session typically ranged from 40 minutes to 1 hour. Up to half the time was spent completing questionnaires. The questionnaires covered the subjects' personality traits, demographics, technical experience, opinions towards a future robot companion, how they felt about the two contrasting robot 'personalities' exhibited by the robot during the interaction scenarios, what they liked or disliked about the robot interactions, and how it could be improved, etc.

6.1 Questionnaire Design

Questionnaire design requires training and experience. There are a number of different considerations researchers should take into account before embarking on designing a questionnaire. Firstly, the notion of whether a questionnaire is the best form of data collection should be addressed. For instance, in some situations an interview format might be preferred (e.g. if conducting robot interactions studies with young children that have low reading abilities). A questionnaire is usually completed by the participant alone. This does not allow the researcher to probe for further information they feel may be relevant to the experiment or verify participant responses. However, the advantage of using questionnaires is that they are usually fast to administer, and can be completed confidentially by the participant.

The development of a questionnaire goes through a series of different cycles. Questions that should be considered are:

- Is the questionnaire I am going to use a valid measure (i.e. does it measure what I really want it to measure)
- Is it reliable (i.e. do I get the same pattern of findings if the questionnaire is administered a few weeks later?),
- Have I used value-laden or suggestive questioning (e.g. "Do you think this robot is humanlike?"), compared to neutrally phrased questions (e.g. "What kind of appearance do you think this robot has?")?,
- Do I want to use a highly structured questionnaire or a semi-structured questionnaire, for example where subjects can express their attitudes towards a particular aspect of the robot interaction in more detail?

Some questionnaires are easier to design than others. For example, a questionnaire that enquires about subject demographics must include items that enquire about age and gender. However, even when considering something as simple as age, the researcher must decide whether to use age categories or simply get the subject to write their age in.

The complexity of questionnaire design occurs when a new research domain is being explored, and human-robot interaction is a perfect example of this. There is no such thing as a perfect questionnaire, but careful team planning and pilot testing can ensure that you have the best possible measure. To carry out a pilot test for a questionnaire, the researcher must recruit independent subjects with the same demographics that they hope to include in the real experiment. Sometimes, it is not easy to get volunteers to participate in a pilot test, but obviously the more responses you get, the more certain you can be of what necessary changes need to be made. It is good practise to carry out the pilot study with approximately 5-10 subjects although this depends on the number of conditions etc in the experiment. In addition to asking the pilot subjects to complete the trial questionnaires, it is recommended to ask them directly whether they found any aspects particularly unclear, complicated or irrelevant etc. One could also ask the subjects whether they would change anything about the overall structure or format, and whether there were important questions that you omitted.

A further issue relates to the type of data you will have to analyse. It is important at the design and pilot testing phase to consider the statistical frameworks that you wish to use, as the questions need to be asked in order to fit their requirements as well as the research goals. For example, continuous scales for questionnaire responses lead to very different analytical frameworks compared to categorical (e.g. yes/no) response formats (i.e. interval versus nominal/ordinal data). Although this process can seem time consuming at the outset, it is certainly worth it, as it is impossible to make changes while the trials are running. An error in the questionnaires could possibly invalidate one or more questions, or in the worst case, the whole questionnaire. As highlighted above, no questionnaire is perfect and we discovered this for ourselves in the adult robot-interaction study. Below we give an example of a possible problematic question and a suggested solution:

Example question
Q. Would you like the robot to approach you at a speed that is?

1) Fast
2) Slow
3) Neither fast nor slow

The above question is phrased in an unspecific way, resulting in, whatever answer is given, little quantifiable information about the preferred approach speed. Due to this lack of a reference point, in practice, most subjects are likely to choose answer 3), as most people want the robot to approach at a speed which is 'just right'. An improved way of asking the question could be:

Suggested improvement to question
Q. Did the robot approach you during the trials at a speed that you consider to be?

1) Too fast
2) Too slow
3) About right

Hopefully, results obtained from this improved question would relate a subject's preferred robot approach speed relatively to the actual speed employed by the robot in a trial. If finer graduations of preferred robot approach speeds are desired, then the trial context and situation must be more closely controlled, with multiple discrete stages, with the robot approaching at different speeds at each stage.

Questionnaire Completion - In our trials it was necessary for some of the questionnaires to be completed in the robot trial area. The subject completed the first questionnaires while the robot was wandering around the trial areas in order to acclimatise the subject to the robot's presence. The two post scenario questionnaires were also administered in the trial area, straight after the respective scenarios, while they were fresh in the subject's memory. We were not able to gain access to the trial area to turn the video cameras off during this time, as we wanted to preserve the illusion that the subject and supervisor were on their own with the robot during the trial. However, there were several other questionnaires and forms, which could have been, administered elsewhere. This would have reduced the amount of video tape used per session. Also the WOZ operators had to sit perfectly still and quiet for the duration of these questionnaires. However, a drawback of administering the questionnaires outside the experimental room is that it changes the context, and might distract the subject etc. Such factors might influence the questionnaire results. Thus, there is a difficult trade-off between savings in recording video tape and other data during the trials, and providing a 'natural' and undisturbed experimental environment.

The environmental context is an important consideration for human-robot interaction studies as questionnaire and interview responses, and observational data will vary depending on the experimental set-up. For example, it would not appear to be problematic to complete a participant demographics questionnaire in the experimental room, which in this case was the simulated living room containing the robot. However, when administering a questionnaire that relates to robot behaviour, appearance, personality or the role of future robot design, the robot and room set-up could influence subject responses. For example, in both the child and adult studies subjects completed a questionnaire at the end of the robot interaction scenarios about their perceptions towards a future robot companion. If the intention is that they consider the robot interaction and robot appearance they have just interacted with in the responses (as it was the case in our study), then this is acceptable. However, the researchers must be aware that subject experiences with the robots in the simulated living room are likely to have influenced their responses in some way.

For trials run in 2004 at the Royal Institute of Technology, Sweden, the WOZ and camera operators were in view of the subject while user trials were taking place [Green et al. 2004]. However, the focus of their study was mainly on human-robot dialogue and understanding, command and control of the robot, which may not have been affected by the presence of other people. We have found that when other people are present, then subjects will tend to interact with those other people, as well as the robot. For our single adult interactions, we wanted to observe the subjects reactions as they interacted only with the robot. Thus, while the experimenter in the adult study stayed in the same room as the subject, she deliberately *withdrew* herself from the experiment by sitting in a chair in a corner and reading a newspaper. Moreover, she did not initiate any communication or interaction with the subjects, apart from situations when she had to explain the experiment or the questionnaires to the subject, or when she had to respond to a verbal query from the subject. We opted for this approach since the study targeted a '*robot in the home'* scenario, where it would be likely that a person and robot would spend a considerable amount of time alone together in the environment.

7 Design and Methodological Considerations

At the outset of designing any study there are a number of crucial design and methodological considerations.

First, the research team must decide what the sample composition will be including, individuals, groups, children, adults, students, or strangers from the street. This is important as the interpretation of results will be influenced by the nature of the sample. For example in the current study, we observed quite distinct differences in the interaction styles between groups of children who were familiar with each other, and individual adults who were alone in the room with an experimenter who did interact in the experiment. Also, as with many other studies, the current adult sample were self-selected and were all based at the university (either as staff or students), which could result in a positive or negative bias in the results. It is very difficult to recruit completely randomised samples and there is always a certain amount of self-selection bias in all studies of this design.

Second, the environmental context should be considered, in the sense of whether a laboratory set-up is used or a more naturalistic field study. Different results are likely to emerge depending on the environment chosen. The adult human-robot interaction study involved a simulated living-room situation within a conference room at the University. Although we tried to ensure it was as realistic as possible, subjects still knew it was not a real living room and were likely to have felt monitored by the situation. Ideally it would be best to carry out future robot-human interaction studies in peoples' homes or work places in order to capture more naturalistic responses and attitudes towards the interactions. However, there are advantages for carrying out laboratory based studies as it allows the researchers greater control and manipulation of potential confounding variables. This cannot be done in the naturalistic field, so it is certainly common practise to begin new research protocols in laboratory set-ups.

Cultural differences are also important if the researchers are hoping for widespread generalisation of the findings. However, this is often impractical, highly expensive and time-consuming.

The overall design of experiments is extremely important in terms of whether between-subject groups (independent measures design) or within-subject groups (repeated measures designs) are used. There are advantages and disadvantages associated with both. Between-subject designs involve different subjects participating in different conditions, whereas within-subject designs mean that the same set of subjects take part in a series of different conditions. Between-subject designs are less susceptible to practice and fatigues effects and are useful when it is impossible for an individual to participate in all experimental conditions. Disadvantages include the expense in terms of time and effort to recruit sufficient participant numbers and insensitivity to experimental conditions. Within-subject designs are desirable when there are sensitive manipulations to experimental conditions. As long as the procedures are counterbalanced, biased data responses should be avoided.

A final consideration should be whether the researchers feel the results are informative based on information recorded at one time point. Human-robot interaction involves habituation effects of some kind and it would be highly useful for researchers to be able to follow-up the same sample of subjects over an extended period of time at regular intervals, to determine whether for example, they become more interested/less interested in the robot, more positive/negative towards the robot and so forth.

Human-robot interaction studies are still a relatively new domain of research and are likely to have a high explorative content during initial studies. It took the science of human psychology many years to build up a solid base of methods, techniques and experience, and this process is still going on at the present. The field of human-robot interactions is still in its infancy and carrying out these initial explorative studies implies that there are not likely to be any concrete hypotheses claiming to predict the direction of findings. This would be impossible at the outset of studies if there are not many previous research findings to base predictions on. The nature of exploratory studies means that there are likely to be many different research questions to be addressed and in any one study, it is simply impossible to consider all possible variables that might influence the findings. However, once exploratory studies have been conducted it should allow the researchers to direct and elucidate more concrete and refined research hypotheses for future, more highly controlled studies.

8 Summary and Conclusions

We have discussed our experiences of running two trials that involved humans and robots physically interacting, and have highlighted the problems encountered.

1. When designing and implementing a trial that involves human and robots interacting physically within the same area, the main priority is the human subject's safety. Physical risk cannot be eliminated altogether, but can be minimised to an acceptable level.
2. There are ethical considerations to be considered. Different countries have differing legal requirements, which must be complied with. The host institution may also have additional requirements, often within a formal policy.
3. Practical ways are suggested in which robots can be programmed or controlled in order to provide intrinsically safe behaviour while carrying out human-robot interaction sessions. This complements work in robotics on developing safe robot motion and navigation planners by other partners within the COGNIRON project and elsewhere [Roy and Thrun, 2002]
4. The advantages of different types of video cameras are discussed, and we suggest that if using network based video cameras, it is wise to use at least one videotape-based camera as a backup in case of network problems, and vice versa. We also suggest some (obvious) ways to optimise camera placement and maximise coverage.
5. Similarly, we suggest it is good practice to have a backup robot available.
6. Sufficient time should be allocated to setup the experimental room and test all equipment and experimental procedures in situ. For example, our study used Radio Frequency (RF) based equipment to monitor and record the comfort level of the human subjects throughout the adult trial. We found that there was interference coming from sources that were only apparent when all the trial equipment was operating simultaneously.
7. Some points to consider when designing questionnaires are made. Completing questionnaires away from the trial area may conserve resources but influence the questionnaire results.
8. A careful consideration of methodological and design issues regarding the preparation of any user study will fundamentally impact any results and conclusions that might be gained.

It is vital that sufficient time is allowed for piloting and testing any planned trials properly in order to identify deficiencies and make improvements before the trials start properly. Full scale pilot studies will expose problems that are not apparent when running individual tests on the experimental equipment and methods. In our own studies the problems we did encounter were not serious enough to damage or invalidate major parts of the trials. We have highlighted other features of our trials we can improve upon, and made suggestions as to how to overcome the problems we have encountered. The lessons learned can be used to improve future trials involving human-robot interaction.

Acknowledgements

The work described in this paper was conducted within the EU Integrated Project COGNIRON ("The Cognitive Robot Companion") and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020. Many thanks to Christina Kaouri, René te Boekhorst, Chrystopher Nehaniv, Iain Werry, David

Lee, Bob Guscott and Markus Finke for their help in designing, implementing and carrying out the trials.

References

ActivMedia Robotics (2005). ActivMedia Robotics Interface for Applications (ARIA). http://robots.activmedia.com/.

BBC Science news (2004). Interactive games run by the COGNIRON team from the University of Hertfordshire at the Science Museum, London.http://news.bbc.co.uk/2/hi/technology/396 269stm 2004.

Brooks, R.A. (1991). Intelligence without representation. *Artificial Intelligence* Vol 47, pp. 139–159.

COGNIRON (2005). Website available at www.cogniron.org. 2005

Crown copyright, (2003). An introduction to health and safety. HSE Books .http://www.hse.gov.uk/pubns/indg259.pdf. ISBN 0 7176 2685 7. 2003

Crown copyright, (1998). Data Protection Act 1998. http://www.hmso.gov.uk/acts/acts1998/19980029.htm#aofs. ISBN 0 10 542998 8. 1998.

P. Dario, E. Guglielmelli, C. Laschi (2001). Humanoids and Personal Robots: Design and Experiments. *Journal of Robotic Systems* 18 (12), pp. 673–690.

T. Fong,, I. Nourbakhsh, K. Dautenhahn (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, Vol. 42, pp. 143-166.

A. Green, H. Hűttenrauch, K Severinson Eklundh (2004). Applying the Wizard of Oz Framework to Cooperative Service discovery and Configuration. Proc. *IEEE Ro-man 2004*, 13th IEEE International Workshop on Robot and Human Interactive, Oka-yama, Japan, IEEE Press.

P. Hinds, T. Roberts, H. Jones (2004). Whose Job Is It Anyway? A Study of Human-Robot Interaction in a Collaborative Task.. *Human-Computer Interaction*, Vol. 19, pp.151-18.

T. Kanda, T Hirano, D Eaton (2004). Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial. *Human-computer Interaction*, Vol. 19, pp. 61-84.

D. Maulsby, S. Greenberg, R Mander (1993). Prototyping an intelligent agent through Wizard of Oz. *Proc.ACM SIGCHI Conference on Human Factors in Computing Systems, Amsterdam*, The Netherlands, ACM Press, pp. 277-284.

B. Robins, K. Dautenhahn, J. Dubowski (2004). Investigating Autistic Children's Attitudes Towards Strangers with the Theatrical Robot - A New Experimental Paradigm in Human-Robot Interaction Studies . *Proc. IEEE Ro-man 2004*, 13th IEEE International Workshop on Robot and Human Interactive. Oka-yama Japan, IEEE Press, pp. 557-562

N. Roy, S. Thrun. Motion planning through policy search (2002). *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, EPFL, Switzerland, pp. 2419–2425.

UPR AS/A/2 (2004). University of Hertfordshire Ethics Policy. http://wisdom.herts.ac.uk/research/Indexes2/usefullinks2.htm. 2004.

VICTEC (2003). VICTEC Project website available at http://www.victec.org, 2003.

Is Someone Watching Me? – Consideration of Social Facilitation Effects in Human-Robot Interaction Experiments

Sarah Woods, Kerstin Dautenhahn, Christina Kaouri

School of Computer Science
University of Hertfordshire
College Lane, Hatfield, AL10 9AB, U.K.
{s.n.woods, k.dautenhahn}@herts.ac.uk

Abstract – **This discussion paper considers the implications of the psychological phenomena of social facilitation effects for human-robot interaction (HRI) studies. Research studies in HRI have significantly increased over the past few years. Such studies investigate e.g. robot appearance and behaviour, and the responses of subjects to robots. However, the possible effects of the experimental context on results from human-robot interaction studies have attracted little attention. In this paper we provide an overview of robot trials with children and adults, as part of the Cogniron project. Observations from video footage are reported, with particular consideration for the influence of the social context and social facilitation effects, including task complexity, evaluation context and type of presence on outcomes of human-robot interaction studies. Lessons learnt and future design implications for human-robot interaction studies are provided.**

Index Terms – human-robot interaction, social facilitation effects, design implications, adult and child trials, video footage.

I. INTRODUCTION

A relatively new area for robotics research is the design of robots that can engage with humans in socially interactive situations. Traditional service robots that are used e.g. to deliver hospital meals or provide security services often require only minimal human-robot interaction [1]. Different examples of robots designed to facilitate robot-human interaction include AIBO [2], Kismet [3], and Feelix [4]. Dario, Guglielmelli and Laschi [5] have provided a useful classification outlining the evolution of robotics. The evolution begins with basic robots within a structured environment such as within industrial automation and manufacturing, to personal robots designed e.g. for use by the disabled and elderly within the home.

Researchers are becoming increasingly interested in many different aspects of human-robot interaction including personality [6-9], speech [10], gestures [11], emotions [4, 12], posture, movement [13], and robot appearance [14-17]. However, the majority of HRI studies to date have adopted a *robot-oriented* approach by focusing on the robot's appearance, behaviour, perceptual abilities, control architecture, interactive skills, and cognitive abilities. One of the goals of our work in the Cogniron project (Cognitive Robot Companion) is to explore the user perspective of robot interaction, therefore adopting a *human-centred* approach. Investigating robot-human interaction from a human-centred perspective involves not only a consideration of the technological requirements of such a robot, but the study of psychological, social and cultural factors, which is a great challenge for HRI robotics research [5, 8, 16, 17].

In this discussion paper we will report on observations made during two different experimental robot trials conducted as part of the Cogniron project investigating different aspects of human-robot interaction. One trial involved *groups* of children and the other experiment involved *individual* adults. These different robotic experimental conditions, have highlighted the complex nature of human behaviour under different situations. Explanations for different interaction styles must take into account age, differences between the participants, and the distinct nature of the tasks that people were involved in. For example, the children's interactions with the robot were within a 'playful' context whereas in contrast, the individual adult trials were under the pretext of 'goal' oriented tasks. However, aside from these differences, we believe the mere presence of other people played a significant role. In this paper, we propose that psychological research findings into the effects of social facilitation could be insightful in explaining some of the human-robot interaction effects recorded. Findings about the mere presence of others in experimental situations could provide some useful guidelines to the wider robotics community for designing and implementing different experimental paradigms. It needs to be emphasised that our trials were *not initially* designed to study social facilitation effects.

The remainder of this paper is structured as follows. First, we will outline the nature of the different robot trials with the children and adult participants. We will then provide some examples of case studies from the child and adult robot trials, and some overall observations that were made. Psychological research findings into the effects of social facilitation will then be presented and the relevance of these findings to the observational data from the human-robot interaction trials will be considered. Lessons learnt from the experiments conducted and future design implications for robot studies involving human subjects will be proposed.

II. METHOD

The human-robot interaction trials

Group-Child interaction study using a play scenario:

In June 2004, 24 exploratory sessions involving 194 children aged 9-11 years (103 boys, 91 girls) were carried out at the University of Hertfordshire. The aim of the study was to consider non-verbal social interactions with regard to spatial distances between a robot and children. Research questions enquired about how the robot could attract children's attention, children's reactions to being the focus of the robot's attention, and how the children's activities were directed towards the robot as opposed to other children. A two by two design, with four experimental conditions was used, a moving/static pointer (selecting children) condition, and a moving/static camera (expressing "attention") condition.

Groups of 10 children participated in the trials and each session consisted of the robot playing two interactive games:

1) *Rotation game*: the robot revolved in the middle of a circle of children, stopped and selected a child by beeping twice and using different pointing conditions. During each round of the game (6 rounds) the selected child was removed from the circle, and the rest of the group moved 0.5m closer to the robot.
2) *Wander game*: The same procedure was used but at each round the remaining children did not move towards the centre and the robot. The robot wandered randomly around the circle of children and then selected a child by facing the child from a distance of 800mm and making 2 beeps.

The trials involved PeopleBot^{TM} robots that were fitted with a lifting arm, which acted as a pointer (Fig. 1). The robot had a small basket holding presents that were presented at each round of the game. Each of the trials was videotaped to allow researchers to analyse more closely the movements, postures and behaviours of the children.

Fig. 1 PeopleBot^{TM} robot with the lifting arm used in the children's robot trials

Questionnaires were administered before and after the children took part in the rotation and wander game (Figs. 2 & 3).

Individual-Adult interaction study using a simulated living room scenario:

This study explored how adults interacted with a single robot in a simulated living room scenario (Fig. 4). Twenty-eight adults recruited from the University of Hertfordshire participated in the robot trials (using a PeopleBot^{TM} robot), which concentrated on a *human-centred* perspective in terms of how the robot's appearance and possible personality attributes were perceived. Levels of comfort with the robot, and the idea of possible robot companions for the home were also considered.

This study was different from the child robot trials, as a more "serious" task-oriented approach was taken. Adults took part in two different tasks, a *negotiated space task* involving moving within the same restricted working area as the robot, and an *assistance task* where the robot interacted with the adult and assisted them with fetching pens. Each adult was exposed to the tasks twice, on one occasion the robot behaved in a 'socially ignorant' manner and in the other instance a 'socially interactive' manner. The order of presentation was counter-balanced and the behaviour styles of the robots were predefined. During the tasks, subjects were asked to solve problems that were meant to be simple and not involve a high cognitive load. Our intention was to provide contexts where the subjects were kept 'busy' (comparable to a person involved in different activities in her home, e.g. reading a book or cooking), but still provide enough opportunity to observe human-robot interactions.

Fig. 2 & 3 Children playing 'rotation' game and 'wander' game

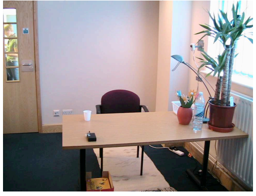


Fig. 4 The simulated living room for the adult interaction trials

Research questions enquired about different situational contexts (cf. Fig. 5), e.g. we investigated which robot behaviour style subjects preferred when the human subject and robot were carrying out different tasks but within the same limited work area.

Adults entered the simulated living room and were introduced to the robot (co-habituation), before completing some initial questionnaires about adult demographics and personality styles. Comfort and social distance tests were conducted between the adult and robot to determine adult comfort levels with the robot. The negotiated space task was carried out with the socially ignorant or socially interactive behaviour of the robot randomised. The assistance task followed with the same robot behaviours. A series of other questionnaires were completed by the adults before they interacted in the same robot tasks again but with the alternate robot social behaviour. At the end of the trials, subjects completed a further questionnaire about their views toward a future robot companion. Results from the questionnaires are presented elsewhere. All trials were videotaped.

Fig. 5 The negotiated space (left) and assistance task in the adult study.

The role of the experimenter

Important differences between the role of the experimenter during the child and adult robot interaction trials should be highlighted. During the children's robot trials, the experimenter dealt with the overall running and management of the game. A guidance role was assumed throughout the trials. They provided instructions to the children about the trials and remained at the periphery of the game to ensure everything ran smoothly. In contrast, the role of the experimenter in the adult trials was more structured. A set of fully standardised instructions was used for each subject. Once the subjects were provided with instructions, the experimenter did not have any eye contact or initiate interaction with them unless directly, verbally, addressed with a question. To maintain this, the experimenter appeared to be reading a large broadsheet newspaper to avoid giving away behavioural cues to the subject.

III. TRIAL OBSERVATIONS

In retrospect, we are aware that the comparison of the group studies with children and individual studies with adults are not strictly comparable due to the different contexts etc. As noted above, the trials were not specifically designed for the study of social facilitation effects. However, the already existing results from these HRI studies served as a starting point for our discussions of social facilitation effects.

Observations from group child-robot interaction trials

Different interaction styles were seen depending on the group set-up, and whether a teacher was present or not in addition to the experimenter. A few selected cases are described below based on observations from the video data.

Child-group robot trials in the presence of an experimenter & teacher - Rotation Game:

When the children were introduced to the robot, they initially sat at quite a distance from the robot, but then two girls moved closer to the centre of the circle towards to robot. The rest of the group then followed. One of the boys then reached out to touch the robot's hand. When the rotation game trial began, all the children tried to have eye contact with the robot when it was pointing with the camera towards them. All the children stood very still with their arms by their sides when the robot was moving around. They looked to each other frequently when the robot initially started moving, and they all giggled when the first boy was chosen by the robot. When the robot started moving again, they became quiet. One of the boys was quite distractible during the game, and put his hands on his hips and seemed to get fed-up waiting for the robot to select a child. The girls tended to smile among each other more than the boys. The closer the children were to the robot, the more they looked at each other, maybe for reassurance of what the robot was doing. When the robot was facing a particular child (camera at the front), the children tended to smile at the robot.

Child-group robot trials in the presence of an experimenter only - Rotation Game:

When this group of children were introduced to the robot, they all sat right in front of the robot (camera and arm side) and started chatting loudly amongst themselves. They seemed to be fascinated with the robot's hand and also wanted to grab the toy 'bugs' from the basket attached to the robot. Some children in the group reached out to touch the robot's hand. This group were much noisier and verbal compared to the group where the teacher was present. In particular, one boy who appeared to act as the group leader influenced the rest of the group. He kept clapping his hands and shouting out very loudly "me, me pick me". This seemed to open up the opportunity for the rest of the group to start saying things to the robot, among "hello, you stupid robot, why won't you pick me?" and "hello, you silly thing". The boy that initiated this dialogue with the robot then started to dance in the hope of getting the robot's attention. When the boy selected by the robot in the first round of the game left the circle he shouted out "I'm dead now". When the next round of the game began the same boy started dancing again and said "If you dance, the robot might pick you". The girls in the group did not seem to pay much attention to this, they stood very still did not verbalise much. The boy who was dancing then moved off the line where the children were supposed to stand, moved toward the robot and shouted out "it wouldn't pick me because of my dance". The boy then flicked the robot's camera. Another boy then started copying the 'dancing' boy and shouted out "oh come on, hurry up and decide".

Child-group robot trials in the presence of an experimenter and teacher - Wander Game:

One of the boys in the group waved at the robot to try and make it start the game. At this point there was a lot of giggling among the group and all the children started looking at each other. There was a lot of eye gaze and attention towards the robot by the children when it was roaming around the room. Some of the children started to wave at the robot to try and get its attention and make it wander towards them. It seemed that two of the boys became somewhat bored with this game and started to look around the room and behind them. There was some chatting among the group. For example, when the robot was slightly delayed in moving, they said 'he's asleep'. When the robot approached one of the girls from the side, it seemed that she was not expecting this and jumped slightly. She then moved the position of her body to face the camera and arm side of the robot.

Child-group robot trials in the presence of an experimenter only - Wander Game:

During the wander game the children were again much louder and more verbal compared to the group where both the teacher and experimenter were present. This group also attempted to gain the robot's attention with different body movements (e.g. waving, dancing) more than the group with the teacher present. At the beginning of this game, one of the children, when it saw the robot approaching another child said "I've got chocolate, come to me". This caused a lot of giggling among the group. Another child began to get frustrated with the robot and shouted out, "Oh come on, it hates me, stupid robot". Another girl when the robot started to approach her said "he likes me". Another child then caught onto the notion that it might influence the robot if it tried to entice it and shouted out "I've got a million pounds robot, come to me, come to me". Children in this group also attempted to name the robot "I name it, Maple school property robot, property of Maple school" and another names it "robot chick". When the robot started to move towards one of the girls that named it, she said "see it answers to its name, I told you that it did".

Observations from individual adult-robot interaction styles

Observations from the video data of individual adult robot trials were completely different from the child robot trials. It was clear from the video footage that the adults were influenced by the presence of the experimenter (e.g. looking towards her for social referencing). The style of adult interactions observed, irrelevant of the robot's behaviour (socially ignorant, socially interactive) ranged from being confident and chatty with the robot to extremely nervous and non-communicative with the robot. Below, we illustrate in more detail the behaviour of five subjects within this spectrum of behavioural responses:

Adult one (female) displayed confident behaviour towards the robot. She chatted with the robot and responded immediately with "hello" when the robot said hello to her. When the subject needed more pens during the assistance task she said "thank you very much" to the robot when they were given to her. This subject had lots of eye contact with the robot and had a relaxed body posture. She also smiled frequently at the robot and said "very nice" and "thank you robot" on a frequent basis. This subject had good task concentration and was able to distribute her attention between the task and robot appropriately (i.e. she wasn't constantly checking to see what the robot was doing, or where the robot was at a cost of completing the task). However, the subject attempted to see how far the robot's abilities extended to and tested to see if it was pre-programmed perhaps, as she initiated conversation with the robot and said "Can you get me a highlighter pen?" The robot then responded with its pre-scripted response "I see that you need some more pens". This subject did not appear to make reference to the experimenter by gazing toward her or asking any questions throughout the trial.

Adult two (male) on the other hand did not engage in any conversation with the robot but did begin to laugh at the robot after a while and smiled occasionally at the robot. This subject appeared to be apprehensive in the presence of the robot to begin with, and was a bit jumpy whilst doing the assistance task as he kept looking up to check on the status of the robot. This subject also checked on the experimenter from time to time, but did not receive any feedback. No questions were directly asked to the experimenter.

Adult three (female) came across as a bit puzzled at the beginning of the assistance task and started to stroke her face gently. She had good eye contact with the robot but at times displayed an unsure smile at the robot. This subject had little conversation with the robot other than saying "thank you" when the robot bought her the pens. She referred to the experimenter during the trial, although not verbally and no feedback was provided.

Participant four (male) appeared to be shocked and puzzled when the robot first spoke to him during the assistance task. He had a very inquisitive, unsure look on his face for quite a while. He hesitated during the task and kept looking up at the robot to see what it was doing. He frequently frowned and raised his eyebrows when the robot made a new noise. He was totally puzzled by the robot's behaviour when the pens were dropped off at the desk. This subject did not have as much eye contact with the robot compared to others and seemed happier ignoring the robot and not paying it any attention. The subject laughed when the robot made a mistake and knocked over the flowers on the desk when dropping off the pens for the task. He raised his cup as a signal of thanks at the end of the trial, but did not attempt to initiate conversation with the robot. When he left the room at the end of the trial, he raised his book, seemingly as a sign of goodbye.

Subject five (female) was the most apprehensive and nervous in the presence of the robot. She jumped when the robot spoke to her and seemed really unhappy and uneasy throughout the trial. She kept making reference to the robot's camera and appeared to be nervous and annoyed when the robot stayed at the desk and didn't move further away. Even though the robot was near to her and had been for a few minutes, this subject jumped with surprise and had a shocked expression on her face when the robot said "I notice that you need more pens". She responded with a nervous smile and continued by biting her lip anxiously.

Based on the above observational data, a number of themes emerged from the trials:

- Overall, children were much more **confident** during the group interactions than adults were during the individual interactions.
- **Conversation** was used more frequently during the group child interactions toward the robot compared to the adult individual trials.
- Children did not seem to make reference to the **experimenter** during the group trials, however, adults frequently checked on the status of the experimenter in the individual trials.
- More **apprehension** and **puzzlement** was exhibited towards the robot in the adult individual trials compared to the group child robot interactions.
- Discrete **body language** and **non-verbal cues** were more evident in the adult trials compared to the child trials.
- Children frequently went right up to the robot to **inspect** it, even when they were not instructed to, and started closely examining the camera and the robot's arm. None of the adults did this with the robot's camera.
- Children assigned a **name** to the robot, and a **gender** on some occasions, and tried to use tactics to attract the robot and get the robot to approach them. No adults exhibited this kind of behaviour.
- The **presence of a teacher** during the child robot trials seemed to have a large effect on the amount of conversation the children exhibited and body movements toward the robot.
- Children in the group situations **copied** and **imitated** each others' actions frequently in an attempt to get the robots attention (e.g. the two boys dancing). Copying of actions of other people did not occur in the adult trials.

IV. SOCIAL FACILITATION EFFECTS

In this section, we provide possible explanations of our observations in relation to social facilitation effects as discussed in the literature. Lessons learnt and implications for HRI studies are presented.

The robot trials described above involved different group configurations ranging from individual adult trials with just the experimenter present, to group child interactions with the experimenter, and sometimes a teacher present. An exploration of psychological research into social facilitation effects in groups could be useful in explaining the different robot-human interactions and the future design of robot experiments paying close attention to the environmental context and group structure. Few studies within robotics research have considered the phenomena of social facilitation. One exception is the study carried out by Bartneck [18] which found that participants gained higher scores in a robot condition compared to a screen condition. Results were explained in terms of social facilitation effects and that the robot character appeared to have stronger social facilitation effect than the screen character resulting in participants putting more effort into the negotiation. In the following sections, social facilitation effects will be explained from a psychological perspective with examples from our human-robot interaction trials. These examples are hoped to highlight the importance of considering social facilitation contextual effects for human-robot interaction studies.

Evidence from Psychology (1)

Social facilitation is one of the oldest social psychology theories in the history of the field of psychology [19]. The theory focuses on changes that occur when individuals perform tasks alone or in the presence of others, and has been defined as the extent to which a given piece of an individual's behaviour is influenced and improved as a result of the real, imagined or implied presence of others [20]. However, social facilitation effects are not straight forward as research is increasingly reporting that the relationship between social presence and individual performance is influenced by: **Task Complexity**, **Evaluation Context** and **Type of Presence**.

In fact, under some circumstances, findings have revealed that performance is inhibited rather than facilitated [19]. Currently, there is still no single theory that can effectively and parsimoniously explain the phenomenon of social facilitation [19]. Below, we provide an overview and description of some of the earliest studies into social facilitation, followed by a summary and critique of different psychological explanations for social facilitation effects.

The social facilitation paradigm is rich in history and dates back to the original experiments on pacing and competition, carried out by Triplett in 1898 [21]. Triplett noted that bicycle racers turned in faster times when they were racing with each other, than when they raced alone. The fastest times were for those cyclists who competed against each other and the slowest times were observed for those who raced against the clock with no pacesetter. In another experiment, Triplett found that most children reeled fishing reel faster when they were reeling alongside another child. Triplett proposed that the presence of a co-actor stimulated a competitive instinct that motivated the individual to reel faster.

However, the experiments conducted by Triplett all contained elements of competition. Allport [22] coined the term 'social facilitation' and attempted to control for competition effects by carrying out experiments using two kinds of mental tasks, word associations and generation of arguments to a written passage. Results revealed that people in group situations made a higher number of associations and generated a larger number of arguments. However, the quality of the arguments generated was better when subjects were alone. This led Allport to suggest that task performance is affected differentially by social presence. Travis (1925) [in 23] proposed 'audience effects' after he observed clear improvements in performance on the pursuit-rotor task, when subjects were confronted with an audience. These findings were supported by Dashiell [24] who reported enhanced performance for simple tasks when an audience was present. However, some researchers found negative audience effects. For example, Pessin (1935) (in [23]) reported that students made more errors in the 'audience' condition when the task was to recall nonsense syllables compared to the 'alone' condition.

Possible Implications for HRI studies based on the Cogniron robot trials (1)

The human-robot interaction trials carried out as part of the Cogniron project emphasised to both the child and adult groups that their performance was not being judged and that there were no correct or incorrect interaction styles. However, based on the video observations, it was clear that some of the children were competing with each other to try and get the robot's attention and get it to select them. In the adult single trials, none of the adults actively tried to get the robot to approach them or instruct it to do something for them. However, indirect elements of competition were also evident from the adult trials. Even though, the activities chosen for the trials (e.g. copying words onto a board and highlighting words) were designed to be non-competitive with no directly observable achievement goals, the adults still behaved as if they were under exam conditions, and took the tasks extremely seriously. The majority of the adults expressed little fun or enjoyment during the tasks, instead expressions of hard concentration were observed, despite the fact that the experimenter emphasised the non-performance based nature of the trials. If a more relaxing, leisurely athmosphere is desirable, then the experimental setup needs to be revised. To assist in obtaining these conditions, a longer habituation time between the adult and robot may be necessary. Alternatively, studies carried out in the field rather than the laboratory could eliminate some of the issues surrounding competition and task performance [25]. For example, robot trials in subjects' homes would be advantageous because although we tried to furnish the conference room at the University to resemble a simulated living room as closely as possible, all the subjects knew that it was not real or natural. Ultimately, adults' focus on task performance could have masked some important findings about their perceptions of the robot, as they may not have observed enough of the robot's behaviour and appearance to form concrete perceptions and attitudes, but more the content of the tasks and their individual performance on it. Future instructions could clearly state that the trials focus on the robot's behaviour rather than the subjects' performance, and that subjects should pay close attention to the appearance and behaviour of the robot interactions.

Evidence from Psychology (2)

Zajonc noted discrepancies in research findings concerning audience and alone conditions, and concluded that well-learned responses are facilitated by the presence of spectators, while the acquisition of new responses is impaired. Therefore, performance is facilitated and learning is impaired by spectators [23]. Zajonc drew a distinction between dominant and non-dominant responses, noting that some behaviours are easier to learn and perform than others. If a task is easy for the person, then the dominant response will be the correct one (i.e. most likely) and therefore the audience/co-actor helps to elicit this. However, sometimes the dominant response is the incorrect one(s) (i.e. the most likely again), but the audience still assists in the elicitation of this response.

Support was generated for the mere presence effect in an experiment carried out by Guerin [26] where participants performed a rotory-pursuit task in one of three conditions: alone, with a simple distraction (a large mirror placed alongside the participants), and in the presence of a (passive) confederate. The confederates did not watch the participants and did not know what the experiment was about. Evaluation effects should not have played a role because the participant could not complete the task and look at the confederate at the same time. Results revealed that subjects performed better when the confederate was present compared to when they were alone or had the distraction of the mirror. Platonia and Moran [27] reported similar results.

Attention has concentrated on the 'alone' condition in some studies, as it has been suggested that this is a poor control for social facilitation experiments. For example Griffin and Kent [28] demonstrated that by simply giving participants a task to perform within laboratory conditions leads them to assume that their performance is being monitored, even without the presence of an audience. Griffin [29] carried out a study with psychology undergraduates involving four conditions: in two conditions an identical card-sorting task was performed either with or without a stopwatch, in the third condition, the card-sorting task was performed in front of an audience of two (one man, one woman), and in the fourth condition, participants waited alone, without a task. Results from questionnaires revealed that participants inferred monitoring in both of the task conditions. Participants who waited alone without a task did not infer monitoring.

Possible Implications for HRI studies based on the Cogniron robot trials (2)

In our robot trials, subjects were told that the experimenter was not part of the task and was not there to monitor individual performance. Although we did not assess the extent in which subjects felt monitored, the video cameras that we used in the adult trials were pointed out at the outset of the trials. Subjects were told that the sole purpose of the cameras was to monitor the robot's behaviour. The extent to which participants felt monitored could be an important factor for future studies to consider as it may influence robot-human interaction styles. This is a difficult situation to overcome for human-robot interaction studies as even without the direct presence of an experimenter, simply giving subjects a task to perform appears likely to elicit some feelings of performance being monitored. The experimental context however could assist in reducing feelings of being monitored. It was clear from the observations of the robot trials that the children in group situations felt much more comfortable and less monitored than the adults did in the alone condition. The nature of the tasks could have had an impact on this also, as the child trials had a much more game-like, relaxed content compared to the adult robot trials. It would be interesting in future adult trials to explore the impact of group size (i.e. pairs, groups of 4) on perceptions of the robot.

Evidence from Psychology (3)

Other studies have considered the effects of group versus alone conditions in relation to impression formation [30]. Thomas et al. carried out an experiment to test whether Zajonc's drive theory of social facilitation also predicted how the mere presence of others influenced social judgement ratings. Participants were exposed to different conditions involving a positive and negative experimenter behaviour condition, and a mere presence condition, group or alone. Results revealed that people formed more extreme impressions of a target when they made judgements in the mere presence of others. This was found for both positive and negative behavioural effects and is consistent with the drive theory of social facilitation.

Possible Implications for HRI studies based on the Cogniron robot trials (3)

Little attention has been paid to the effects of an experimenter in human-robot interaction studies, and may imply confounding variables. For example, in the current robot experiments, it could be the case that the children formed more extreme positive and negative opinions toward the robot under the group condition compared to the adults who performed the interactions alone. Children in groups were able to discuss between themselves what they felt towards the robot and verbalised frequently about 'the robot being stupid because it didn't approach them' or 'that the robot was clever because it came to them'. Discussing opinions within a group situation is an important aspect of impression formation. However, the adults did not have this opportunity and were unable to confirm or disconfirm their attitudes and perceptions towards the robot with anyone, which could account for the frequent appearances of puzzlement on subjects' faces. The same experimenter was present at both the child and adult robot trials, and never exhibited negative behaviours, so this was unlikely to have had an impact on the robot trial results.

A further important consideration is whether the presence of others facilitates or inhibits emotional expression. For example, in our human-robot interaction trials, we are interested in capturing the emotional responses and reactions that participants displayed in response to the different robot scenarios. However, it is possible that the video footage of emotional responses was affected by the group and individual conditions, meaning that direct comparisons should not be made. Ross et al. [31] carried out a study to examine the potential effects of group structure on emotional responses. Participants viewed emotionally loaded slides with either friends, strangers or alone. Results revealed that strangers had overall inhibitory effects on communication accuracy, whereas friends had facilitative effects on slides showing positive emotions and inhibitory effect on emotionally negative slides. Results for the alone condition were somewhere in between the stranger and friend conditions. These effects could have been evident during our robot trials. For example, the children were in the presence of their classmates and friends, which could have facilitated the expression of different emotions during the robot trials. In contrast, an inhibitory effect on emotion expression may have occurred with the adult trials as they might have tried to conceal their emotions as they felt more self conscious not having anyone to confirm or discuss their feelings with.

Theoretical explanations for social facilitation affects

Guerin [32] classified social facilitation effects according to three theoretical perspectives, drive theories, social comparison theories, and cognitive process theories. Drive theories were inspired by Zajonc [23] and were based on the Hull-Spence drive theory, which posited that in the presence of others, individual drive levels are elevated. When arousal arises from a difficult or unfamiliar task, this results in stress and consequently poor performance. This extra arousal results in taking people past the optimum arousal level and results in the dominant response being elicited whether it is easy, or new and difficult. Zajonc was challenged on his ‘mere presence’ explanation, but later asserted that social facilitation effects still emerged even when the situation and behaviour of others were controlled for. For example, factors such as evaluation apprehension could influence individual reactions to the presence of others, but Zajonc still claimed that mere presence was necessary and sufficient for social facilitation. Cottrell [33] was one of the proponents that challenged Zajonc’s drive theory, and stated that mere presence was not enough to elevate drive levels and would not necessarily cause social facilitation effects. Cottrell proposed the evaluation apprehension effects hypothesis stating that increases in drive levels were a result of individuals being concerned about how others would evaluate them. Cottrell also stated that prior evaluation experiences caused people to develop a drive reaction – a learned drive.

An example of a social comparison theory is the Self-presentation explanation. Self-presentation assumes that social facilitation effects are focused on impression formation and the well-established finding that people are motivated to please those that are observing them, sometimes referred to as social desirability effects. Baumeister [in 19] for example suggested that the presence of someone considered to be evaluative would trigger more drive than the presence of someone not evaluating performance.

Cognitive processes as an explanation for social facilitation effects often emphasise distraction (in [19]). For example, attention conflict (Baron 1986, in [19]) can produce drive like effects on performance that can facilitate simple tasks and impair complex ones.

Future directions for the design of social facilitation studies and the implications for human-robot interaction research

Aiello and Douthitt [19] highlighted some of the issues that future research studies into social facilitation research should consider. Firstly, studies should consider the nature of the group composition and levels of familiarity. For example, effects might be different for friends, family members, coaches, teachers, supervisors etc. Most studies to date have been carried out with groups of participants who are assumed to be strangers. However, the child robot-interaction studies in the current human-robot interaction trials were carried out with groups of children from the same class, who were highly familiar and possibly friends with each other, and in some cases with the teacher present. In comparison, the adult robot-interaction trials were conducted with individual adults and just the experimenter who was assumed to be a stranger in most instances. The effects of group structure and familiarity could therefore have had a considerable impact on the results of subjects’ perceptions toward the robot. Second, studies should explore the role of time and how long the predicted effects of social facilitation are expected to last. The practical implications for social facilitation effects would be different if they lasted over days or weeks, compared to a few hours. This is an important point to take into account for the robot trials as habituation effects with the task set-up and robot could greatly influence results. Tasks employed to study social facilitation effects have relied on the quantity and quality domains and have not taken into account contextual performance issues. Finally, measures used to explain social facilitation effects should be improved and should not rely too heavily on self-report assessments or through inferences from manipulating experimental conditions (e.g. condition where people are told they are being evaluated or not evaluated). Although, our human-robot interaction trials were not designed for measuring social facilitation effects, the importance of considering the effects of monitoring, group formation, social context, task complexity and the methods used to assess subjects perceptions towards robots should be carefully planned and piloted.

V. CONCLUSIONS

To summarise, we have presented observations from video footage data of human-robot interaction trials, and discussed these in light of psychological findings related to social facilitation effects. The design of research studies into human-robot interaction is growing. However, the possible effects of the *social context* where experiments take place is usually not addressed, although awareness of these effects is beginning to emerge. For example, a longitudinal case study trial using the service robot Cero emphasised the importance of considering the situational context when analysing and interpreting results and conclusions [25]. It can be expected that the social context of any HRI study will impact the results. This context will vary considerably depending on the scenario used, the type of robot, and desired purpose for the robot. For example in the case of ‘toy’ robots such as AIBO

[2], and Kismet [3], where the only purpose of the robot is to engage people in interaction, the social context of human-robot interaction is likely to be very different compared to adult interactions in a more 'serious' task context involving service robots (e.g. [10, 25, 34, 35]). Also, the social context of a robot that serves as a museum guide or is shown at exhibitions is likely to impact how people will behave in the robot's presence. An observation of distinct differences between our child and adult robot trials under different social contexts resulted in a discussion of the relationship between social presence and individual performance in relation to task complexity, evaluation contexts and type of presence. The importance of acknowledging these social effects in future studies is emphasised, as ignoring these influences could significantly impact on findings, and ultimately the design implications for future robot companions. While the examples of human-robot interaction discussed in this paper were not specifically designed for the study of social facilitation effects, future work can take into account the issues that we raised.

ACKNOWLEDGMENTS

The work described in this paper was conducted within the EU Integrated Project COGNIRON ("The Cognitive Robot Companion") and was funded by the European Commission Division FP6-IST Future & Emerging Technologies under Contract FP6-002020. Many thanks to Mick Walters, Kheng Lee Koay, Christina Kaouri, René te Boekhorst, Chrystopher Nehaniv, Iain Werry, & David Lee for their help in designing, implementing & carrying out the trials.

REFERENCES

[1] D. Wilkes et al., "Toward socially intelligent service robots," *Applied Artificial Intelligence Journal*, vol. 12, pp. 729-766, 1997.

[2] Sony, "Sony Entertainment Robot Europe http://www.aibo-europe.com/," 2004.

[3] C. L. Breazeal, *Designing sociable robots*. Massachusetts: The MIT Press, 2002.

[4] L. Cañamero, "Playing the emotion game with Feelix: What can a LEGO robot tell us about emotion?" in *Socially intelligent agents: Creating relationships with computers and robots*, K. Dautenhahn, A. Bond, L. Canamero, and B. Edmonds, Eds. Massachusetts, USA: Kluwer Academic Publishers, 2002, pp. 69-76.

[5] P. Dario, E. Guglieimelli, and C. Laschi, "Humanoids and personal robots: Design and experiments," *Journal of Robotic Systems*, vol. 18, pp. 673-690, 2001.

[6] D. C. Dryer, "Getting personal with computers: How to design personalities for agents," *Applied Artificial Intelligence*, vol. 13, pp. 273-295, 1999.

[7] K. Dautenhahn, "Design spaces and niche spaces of believable social robots," *Proc. IEEE Ro-man*, Int. Workshop on Robots and Human Interactive Communication; RoMan, Berlin, Germany, 2002.

[8] S. Woods, K. Dautenhahn, and J. Schulz, "The design space of robots: Investigating children's views," *Proc. IEEE Ro-Man*, Kurashiki, Japan, pp. 47-52, 2004.

[9] K. Severinson-Eklundh, A. Green, and H. Hüttenrauch, "Social and collaborative aspects of interaction with a service robot," *Robotics and Autonomous Systems*, vol. 42, pp. 223-234, 2003.

[10] B. A. Maxwell et al., "Alfred: The robot waiter who remembers you." *Proc. 8th AAAI Workshop*, Texas, USA, 2000.

[11] M. Scheeff et al., "Experiences with Sparky, a social robot," in *Socially intelligent agents: Creating relationships with computers and robots*, K. Dautenhahn, A. Bond, L. Cañamero, and B. Edmonds, Eds. Massachusetts, USA: Kluwer Academic Publishers, 2002, pp. 173-180.

[12] R. C. Arkin, *Behaviour based robotics*. Massachusetts: The MIT Press, 1998.

[13] M. Montemerlo et al. "Experiences with a mobile robotic guide for the elderly," *Proc. AAAI National Conference on Artificial Intelligence*, Edmonton, Alberta, Canada, 2002.

[14] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and Autonomous Systems*, vol. 42, pp. 143-166, 2003.

[15] B. R. Duffy, "Anthropomorphism and the social robot," *Robotics and Autonomous Systems*, vol. 42, pp. 177-190, 2003.

[16] C. DiSalvo et al., "All robots are not created equal: The design and perception of humanoid robot heads," *Proc. DIS2002*, London, 2002.

[17] J. Goetz, S. Kiesler, and A. Powers, "Matching robot appearance and behaviour to tasks to improve human-robot cooperation," *Proc. IEEE Ro-man,* Milbrae, CA, 2003.

[18] C. Bartneck, "Interacting with an embodied emotional character," presented at DPPI, Pittburgh, Pennsylvania, USA, 2003.

[19] J. R. Aiello and E. A. Douthitt, "Social facilitation from Triplett to electronic performance monitoring," *Group Dynamics*, vol. 5, pp. 163-180, 2001.

[20] D. Forsyth, *Group Dynamics*. New York: Brooks/Cole-Wadsworth, 1999.

[21] N. Triplett, "The dynamogenic factors in pacemaking and competition," *American Journal of Psychology*, vol. 9, pp. 507-533, 1898.

[22] F. H. Allport, "The influence of the group upon association and thought," *Journal of Experimental Psychology*, vol. 3, pp. 159-182, 1920.

[23] R. B. Zajonc, "Social faciliation: A solution is suggested for an old unresolved social psychological problem," *Science*, vol. 149, pp. 269-274, 1965.

[24] J. F. Dashiell, "Experimental studies of the influence of social situations on the behaviour of individual human adults," in *Handbook of social psychology*, C. A. Murchison, Ed. Worcester MA: Clark University Press, 1935, pp. 1097-1158.

[25] H. Hüttenrauch and K. Severinson-Eklundh, "Fetch-and-carry with CERO: Observations from a long-term user study with a service robot," *Proc. IEEE Ro-man 2002*, Int. Workshop on Robot & Human Interactive Communication, 2002.

[26] B. Guerin, "The effects of mere presence on a motor task," *The Journal of Social Psychology*, vol. 126, pp. 399-401, 1986.

[27] J. Platonia and G. Moran, "Social facilitation as a function of the mere presence of others," *The Journal of Social Psychology*, vol. 14, pp. 190-197, 2001.

[28] M. Griffin and M. Kent, "The phenomenology of the alone condition. The role of aloneness in social facilitiation," *The Journal of Social Psychology*, vol. 138, pp. 667-669, 1998.

[29] M. Griffin, "The phenomenology of the alone condition: More evidence for the role of aloneness in social facilitation," *The Journal of Psychology*, vol. 135, pp. 125-127, 2001.

[30] S. L. Thomas et al., "Social facilitation and impression formation," *Basic and Applied Social Psychology*, vol. 24, pp. 67-70, 2002.

[31] B. Ross et al., "Social facilitation and inhibition of emotional expression and communication," *Journal of Personality and Social Psychology*, vol. 63, pp. 962-968, 1992.

[32] B. Guerin, *Social Facilitation*. Cambridge, England: Cambridge University Press, 1993.

[33] N. B. Cottrell, "Social facilitation," in *Experimental social psychology*, C. G. McClintock, Ed. New York: Holt, 1972, pp. 185-236.

[34] S. Thrun, et al., "Probablistic algorithms & the interactive museum tour-guide robot Minerva," *Journal of Robotics Research*, vol. 19, pp. 972-1000, 2000.

[35] H. Hüttenrauch & K. Severinson-Eklundh, "To help or not to help a service robot," *Proc. IEEE Ro-man 2003*, Int. Workshop on Robot & Human Interactive Communication, Millbrae, CA, USA, 2003.

Comparing Human Robot Interaction Scenarios Using Live and Video Based Methods: Towards a Novel Methodological Approach

Sarah Woods, Michael Walters, Kheng Lee Koay, Kerstin Dautenhahn
Adaptive Systems Research Group, School of Computer Science
University of Hertfordshire
College Lane
Hatfield, UK, AL10 9AB
{S. N. Woods, M.L.Walters, K.L.Koay, K.Dautenhahn} @herts.ac.uk

Abstract — **This paper presents results of a pilot study that investigated whether people's perceptions from live and video HRI trials were comparable. Subjects participated in a live HRI trial and videotaped HRI trials in which the scenario for both trials was identical, and involved a robot fetching an object using different approach directions. Results of the trials indicated moderate to high levels of agreement for subjects' preferences, and opinions for both the live and video based HRI trials. This methodology is in its infancy and should not be seen as a replacement for live trials. However, our results indicate that for certain HRI scenarios videotaped trials do have potential as a technique for prototyping, testing, developing HRI scenarios, and testing methodologies for use in definitive live trials.**

I. INTRODUCTION

In the course of our work for the COGNIRON Project [2005], we are primarily interested in the research area of Human-Robot Interaction (HRI), in particular with regard to socially interactive robots. An excellent overview of socially interactive robots is provided in Fong et al.[1]. We are primarily interested in the human perspective of how robots could be useful in domestic environments; in particular the roles, tasks, and social behaviour that will be necessary for robots to exhibit in order to integrate into normal domestic situations. In order to study human-robot relationships, we typically run HRI trials using carefully devised test scenarios, where human responses and opinions can be collected using a variety of methods. HRI trials are particularly difficult to develop if they should involve complex robot behaviours that nevertheless need to be reliable and replicable for statistical comparisons, various robot platforms etc, in addition to large sample sizes of subjects, balanced for age, gender, cultural background etc. In order to address this situation, we are interested in verifying whether videotaped HRI trials for various scenarios could be used in certain situations instead of live HRI trials, or as a complementary methodology for live trials.

A. Human Robot Interaction Trials

To date, we have conducted various live HRI trials with human scaled PeopleBot^{TM} robots [2][3][4] Other researchers have also conducted similar HRI trials with human sized robots including Dario et al. [1], Severinson-Eklundh et al.[5], Kanda et al.[6] and Hinds et al.[9].

However, most HRI trials to date typically are characterised by relatively small sample sizes [6][7][8]. Our largest HRI study to date involved 28 subjects and took about 2 months to carry out. We have run several larger HRI trials, but this has been at the expense of compromising the trial conditions (audience noise, lack of extensive post trial questions etc). Running a live HRI trial under controlled conditions therefore requires a major commitment of time, resources and personnel to ensure that statistically valid results are obtained. HRI studies in general are at a stage where there is not a large body of prior work to guide the design of large scale live trials. This means that most studies are highly exploratory. Many initial assumptions are based on those expected from human-human interactions, which we and others have found do not always hold true for human-robot interactions [9][10][11][12]. It is sometimes difficult to justify speculative or exploratory trials where, by their nature, there is a higher probability of the predicted assumptions not being met, being inconclusive or irrelevant. Before committing to a major trial it is essential to run pilot studies to test the proposed methodology. It would be advantageous to have a methodology in place where trial predictions could be piloted and tested, before developing and executing full live trials.

B. 1.2 Video Based HRI Trials

To overcome some of the drawbacks of live HRI trials, the feasibility of running HRI trials using video footage rather than a full live interaction was considered. Although this methodology would certainly be inferior to a live HRI session, it was hoped that it would yield valuable results towards the development of live trials. Kidd [13] found no significant differences between subjects' ratings of personality traits for 'present' and 'remote' (through video) cases of an interaction with a robot head. Shinozawa et al. [14] reported that comparing a robot's recommendation behaviour with an on-screen agent's, for human decision making, depended on the interaction environment and that geometrical consistency between the interaction environment, and robots and on-screen agents was important. Paiva et al. [15] reported that synthetic (cartoon-like) characters in virtual environments were readily empathized with by children as they enacted various scenarios. This provides supporting evidence that believable relationships can be created through the medium of video. Using videos of robots, which are more realistic than virtual or synthetic characters, could result in HRI trials that are even closer to resembling real live interactions. Video based HRI trials have the potential advantages to: 1) reach larger numbers of subjects as they are quicker to administer, 2) easily incorporate subjects' ideas and views into later video trials simply by recording extra or replacement scenes into the video based scenarios, 3) carry out trials exposing groups of subjects to the HRI scenario simultaneously, 4) prototype proposed live trial scenarios to avoid wasted effort and test initial assumptions, 5) allow greater control for standardised methodologies (i.e. exactly the same robot behaviours, exact trial instructions etc.). As this is an unexplored area of HRI studies, it is first necessary to confirm whether video based HRI trials are able to provide comparable results to live trials, and also under what circumstances. A pilot study using both live and video HRI trials was developed to begin exploring the following main research questions:

1) Will video based HRI scenarios provide results that are comparable to results obtained from live HRI trials?

2) Under what circumstances would video based trials provide comparable results to live HRI trials?

3) What are the likely limitations of video based trials in gaining valid human responses to HRI scenarios?

C. *Experimental Method*

The Video-Live Trial (VLT) pilot study was carried out in a converted conference room. The chosen scenario involved a robot using different approach directions to bring a seated subject an object. The aims of the trial were to find out about subject preferences for the robot approach directions. The room was partitioned into two areas; a video trial area and a live trial area. There was a gap in the partition, so that it was possible to move between the two areas (see Fig.1) but not possible for subjects to see the other area while carrying out the respective video or live trials. The live trial area resembled a simulated living room with a chair and two tables. The subject was seated in the chair throughout the live trial which was positioned halfway along the rear wall (point (9), Fig.1). To the left front and right front of the chair, two tables were arranged (with room for the robot to pass by) in front of the chair. One of the tables had a television placed upon it; the other had a radio and CD player.

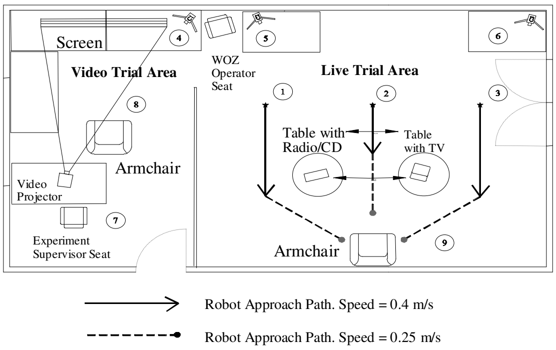


Figure 1. Diagram of video and live trial experiment areas

The robot was driven to the appropriate start position by an operator seated at a table in the far corner of the room. Subjects were told that the robot would be controlled by the operator while it was driven to the three start positions, but would be approaching them autonomously to bring them the TV remote control. This was reinforced as the operator made notes and did not press any of the robot control keys (on the robot control laptop) while it approached the subject.

The video trial area contained a video projection screen and projector for playing the video HRI trial scenarios. The videos were all recorded in the live trial area, with an actor playing the part of the subject. The actor was male, and the narration voice which introduced and set the scene for the HRI trial scenario was also male. The videos were recorded using a mixture of first and third person points of view. The third person views showed the overall positions and actions of both robot and (actor) subject. Then by switching to a first person view (from the perspective of the subject sitting in the chair as the robot approached) a viewer saw the robot approaching in a way that was as realistic as possible and could gain some spatial perspective (see Fig. 2 for example screen shots).

D. *The HRI Trial Scenario*

An identical scenario was used for both the video and live HRI trials and took place in a (simulated) living room (Fig. 1). It was introduced either by the experiment supervisor for the live trial, or by the narrator for the video based trial. The context was that the subject had arrived home from work and rested in an armchair (point (9), Fig.1). The subject then asked the robot to fetch the remote control. It was explained to the subject that the robot was new to the household and it was necessary to find out which approach direction the subject preferred; either from the front (2), the left (1) or the right (3) (see Fig. 1). In order to justify the robot fetching the remote control, one of the tables had a (switched off) TV set upon it. The other table had a CD-Radio unit. Our expectations prior to the trials were that subjects would prefer the approach from the front, since the robot was then fully visible at all times.

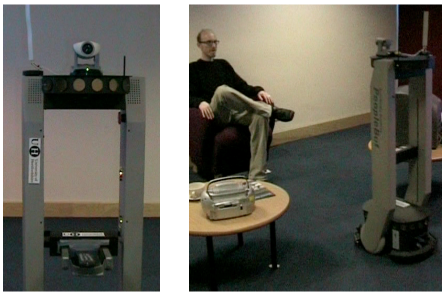


Figure 2. Examples of first and third person views

E. *Experimental Conditions*

We were aware from a previous demonstration that the TV was a natural focus of subjects' attention and could have influenced the choice of preferred robot approach direction. Therefore, half the trials (for both live and video versions) were carried out with the TV on the left hand table, and the other half with the TV on the right hand table.

Each subject experienced the robot approaching from three directions: front, left and right in a counterbalanced order sequence covering all six possible permutations of the three robot approach directions. This was used for both video and live trials. As a consistency check, the three robot approach directions were also repeated (in a different order) for each trial. In order to counterbalance for effects due to the order in which subjects experienced the video and live trials, we exposed half the subjects to the live trial first, then vice versa for the other half of the trials. Fifteen subjects (9 (60%) males; 6 (40%) females) individually participated in the study. The mean age of the sample was 33 years (range 21-56 yrs). Only one subject was left handed. Four subjects were secretarial staff from the University of Hertfordshire, 5 subjects were MSc students studying 'Artificial Intelligence', and the remaining 6 were research staff in the Computer Science Department at the University.

F. 2.5 Procedure

A short introductory questionnaire was used to gain the necessary demographic and personal details from the subjects. At the end of each video or live HRI trial a short questionnaire was used to assess the subjects' views on approach direction, approach speed, stopping distances, comfort levels and practicality for the different approach directions. After both video and live trials had been completed, subjects participated in a semi-structured interview with a psychologist. The interview was carefully designed so that no leading questions were asked. The interviewer was able to follow up answers to gain a deeper insight when necessary. The main purpose of the structured interview was to assess the subjects' views on the trial procedures and methodology, establish any weaknesses and find out how the trial could be improved from the participants' point of view. The subjects' reactions to both live and video based HRI trials were recorded on video tape.

II. RESULTS

A. Approach Direction most preferred and least preferred.

Results of the trials clearly demonstrated that the least preferred approach direction was the front approach, for both the live and video trials. The right approach direction was the most preferred for both the live and video trials, and the left approach direction was preferred equivalently for both live and video trials. Only one person could not state a preference for any of the approach directions based on the video data. Fig. 3 illustrates that there was approximately 58% agreement for the different approach direction preferences, between the live and video trials. 36% (N: 5) of subjects stated that they preferred the right approach direction in both trials, and 22% (N: 3) rated that they preferred the left approach in both trials. Surprisingly, no subjects preferred the front approach direction in both trials. Where agreement was not found between the live and video trials, this was predominantly found for changes in preferences for the left and right approach direction. One subject (7%) stated that they preferred the front approach for the live trial, but the left approach direction for the video trial. One subject (7%) preferred the left approach direction for the live trial, but the front approach direction for the video trial. Two subjects (14%) preferred the left approach direction in the live trial, but the right approach direction in the video trials. Finally, 2 subjects (14%) preferred the right approach direction for the live trial, but the left approach direction for the video trial.

Cross-tabulation tables were produced to calculate the percentage of agreement between subjects' approach direction preferences for the live versus video trials. McNemar-Bowker test, a nonparametric test was carried out to detect changes in responses between the live and video trials. No significant effect was found indicating that there were no significant differences in subjects' approach direction preferences in the live and video methods used.

Fig. 4 demonstrates that there was 85% overall agreement between subjects' ratings of the approach direction they least preferred for the live and video trials. 77% of subjects stated that they least preferred the robot front approach direction in both the live and video trials. One subject (7.7%) least preferred the left approach direction in both trials. No subject rated the right approach direction as least preferred in either the live or video trial.

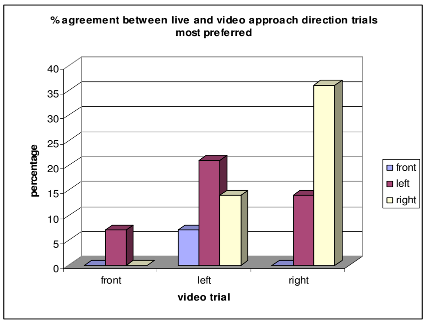


Figure 3. Percentage agreement between approach direction most preferred for the live & video trials

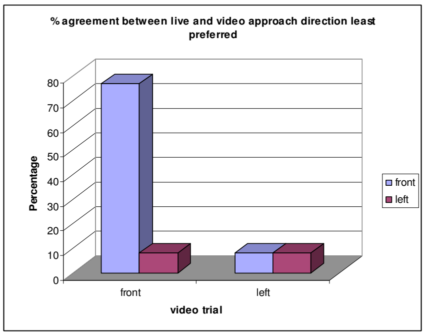


Figure 4. Percentage agreement between approach direction least preferred for the live & video trials

Only one subject (7.7%) who least preferred the front approach in the live trial, had rated the left approach direction as least preferred in the video trials. One subject (7.7%) least preferred the left approach direction in the live trial, but the front approach direction in the video trial. A McNemar test revealed no significant differences in the approach direction least preferred by subjects in the live and video method robot trials.

B. Robot Stopping Distances

1) Front Approach Stopping Distance;

The robot was set up to stop automatically at a distance of approximately 0.50m from the leading edge of the robot and the nearest part of the person's body. There was some overshoot or undershoot giving a tolerance of +/-0.15m. Overall, there was 66% agreement between subjects for ratings of whether the robot's stopping distance was too close, about right, or too far for the live and video trials (Fig. 5).

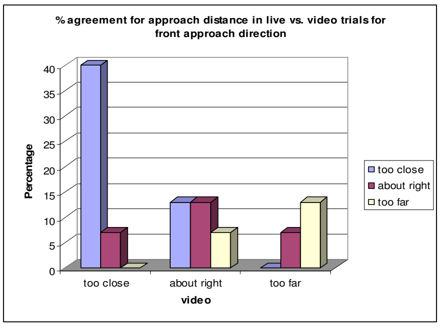


Figure 5. Percentage agreement for robot stopping distance from subject for front approach direction

Figure 5 indicates that a majority of subjects rated the front approach stopping distance as too close. For subjects who rated the front stopping distance as being too far, we observed that these subjects usually had their legs stretched out in front of them. This caused the robot to stop when it reached the subject's feet, rather than moving close enough for them to reach the TV remote control (due to the robot's stopping safety mechanism). Overall, there was 60% agreement between the live and video trials for subjects' ratings of the robot's stopping distance for the left approach direction. There was overall 80% agreement for the robot's stopping distance for the right approach direction in both the live and video trials.

C. *Robot's Speed during the Live and Video Trials*

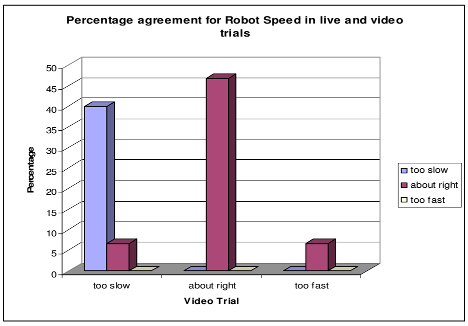


Figure 6. Percentage agreement for robot speed in live and video HRI trials.

The robot used two speeds; the normal approach speed was 0.4m/sec, and the robot slowed to 0.25m/sec when it was within 1m of the front of the subject. When subjects were asked to rate the speed of the robot's approach direction, there was high overall percentage agreement between the live and video trials (87%). 46% (N: 7) of participants rated the robot's speed as about right for both the video and live trials, and 40% (N: 6) of subjects rated that the robot's speed was too slow in both the live and video trials. Where agreement was not found, one subject (6.7%) rated the robot speed during the live trial as about right, but too slow for the video trial, and one subject (6.7%) rated the robot speed in the live trial as about right, but too fast in the video trial. No subjects rated the robot speed as too fast during the live trials (Fig. 6).

D. *Practicality and Comfort for the different Robot Approach Directions*

In addition to subjects rating which approach direction they preferred for the live and video robot trials, ratings were given for how 'practical' they thought each approach direction was for the given task of delivering a TV remote control, according to a 5-point Likert scale (1 = not practical at all to 5 = very practical) (Fig. 7). Paired samples t-tests did not reveal any significant differences between subjects' ratings of practicality for any of the approach directions between the live and video trials [left approach direction t = 1.47 (14), p = .16, right approach direction t = .521 (14), p = .61, front approach direction t = -1.08 (14), p = .30]. This indicates that there were no large discrepancies in subjects' judgments and subsequent ratings of task practicality between the live and video trials. Ratings of the practicability of the approach directions for both the live and video trials were relatively high. The front approach direction received the lowest ratings of practicality for both the live and video trials. The left approach direction in the live trial received the highest ratings of practicality followed by the right approach for both the live and video trials

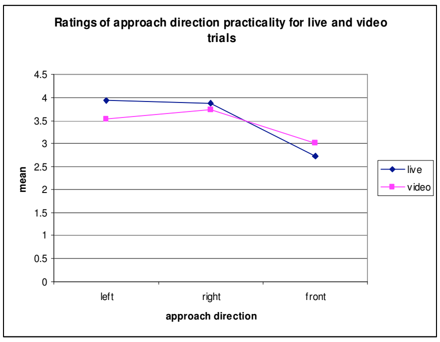


Figure 7. Ratings of approach direction practicality for live and video HRI trials

Subjects were asked to rate how comfortable they felt with the different approach directions for both the live and video trials according to a 5-point Likert scale (1 = very uncomfortable, 5 = very comfortable). Paired samples t-tests revealed no significant differences between subjects' comfort level ratings according to the different approach directions between the live and video trials [left approach direction t =1.23 (14), p = .24, right approach direction t = 1.58 (14), p = .14, front approach direction t = -.52 (14) p = .61]. This highlights that subjects' comfort ratings were relatively equivalent for both the live and video trials. The lowest mean comfort levels were found for the front robot approach direction for both the live and video trial. The highest comfort level rating was found for the right approach direction in the live trial followed by left approach direction in the live trial. It seems that participating in the live trial led to slightly higher (although non-significant) ratings of comfort compared to the video trials. No significant differences were found between the most preferred approach direction and least preferred approach direction for gender, subject handedness (whether subject was left or right handed), and occupation.

E. Realism of the Live and Video Robot Trials

Subjects were asked to rate the overall realism of the live and video robot approach direction trials ranging from 1 = not realistic to 5 = very realistic. The mean rating was 3.47 (minimum = 2, maximum = 4). Subjects were subsequently asked how they would have improved the trials. The comments that subjects made were classified according to those that referred to the set-up of the trial and room, and those that referred to the robot's characteristics:

Subjects' suggestions about how the trials could be improved included a more chaotic and messy environment, more objects and obstacles (as the room was too sterile and clinical), more furniture, and a more busy environment. They would also like the robot to indicate that it understands the person (e.g. by eye gaze), be smaller (as less threatening) and also fetch the remote control itself at start of task (the robot had it in the basket already).

93% of subjects stated that they preferred the live robot trials compared to the video trials which was not surprising. However, 80% of the sample stated that they felt the video robot trials were representative of the live robot trials.

III. DISCUSSION

The main findings from this study were:

1) The level of agreement between subject responses for the preferred robot approach direction was relatively high (60%) between the live and video trials. Discrepancies were mainly due to the fact that subjects did not have strong preferences for either the left or right robot approach direction and sometimes changed these preferences between the video and live trials.

2) Very high levels of correspondence (85%) were found for subjects least preferring the front robot approach direction in both the live and video trials.

3) Moderate to high levels (60-80%) of agreement were found for perceptions of the robot's stopping distance from the subject, for each approach direction in the live and video trials.

4) High agreement (87%) was found for subject ratings of the robot's speed between the live and video trials.

5) No significant differences were revealed between subject ratings of how practical and comfortable the different robot approach directions were for both the live and video trials.

6) Subject ratings for the realism of the video trials in comparison to the live trials were moderately high, although 93% stated that they preferred interacting in the live trials.

These results support findings from an informal earlier study, and also provide additional support for using video methods. We had thought that subjects might find it difficult to perceive the robot distances, and speed through the video medium, but this may not be the case. The non-significant findings for subject ratings of the practicality for the robot approach direction task, and comfort levels between the live and video trials was positive. This is indicative of subjects being able to report on the subjective experience of how comfortable they would feel with different robot approach directions through video footage.

Most subjects preferred the live robot-interaction trials. This was not surprising as live trials seem more interactive[1], likely to be more fun, and more engaging, compared to watching the interactions involving a stranger on a screen. The embodiment experience of being part of a live-set up is also likely to be much more beneficial for assisting in the perception of speed, distances and different robot movements compared to video footage. However, most also reported that the video robot trials were representative of the live trials. Subjects' overall ratings of the 'realism' of the approach direction robot trials was moderately high and most of the improvements that subjects cited were related to the environmental set-up, and context, rather than characteristics of the robot. According to our subjects' suggestions, our future robot trials should take place in a more naturalistic 'messy' living room set up, which is more representative of a realistic home environment.

The results of this pilot study are in line with the findings reported by Kidd [13]. We are not aware of any further studies to date that have considered the suitability of using video footage for human-robot interaction studies. Naturally, there are numerous limitations of using video footage for HRI studies, and we are by no means suggesting that they should be a replacement for live HRI studies. It could be that the more interaction between a robot and a subject in a trial, the less suitable video trials will be. The timing and synchronization of movements play an important part in regulating and sustaining meaningful human-human interactions. Developmental psychologists (e.g. [16][17]) have shown that while babies happily interact with their mothers via live video, they get highly distressed when watching pre-recorded or replayed videos of their mothers (as it lacks the contingency between mother's and baby's behaviour). However, for the particular research questions that we consider in the context of robot motion planning and approach directions, contingency of robot and human movements plays a less crucial role and thus lend themselves to investigations of video trials.

Only 15 subjects participated, many with a robotics or computer science related background that may have biased the results. However, such subjects are most likely to be future customers of a robot assistant in the home. Nevertheless, naïve subjects who have no prior experience with robots might form an interesting control group in future studies. We limited the sample size for this initial pilot study but given the current positive findings, we aim to replicate the current study with a larger sample size. In future trials, we also intend to incorporate more naturalistic set-ups. The quality of the robot trials could be enhanced if professional camera techniques are adopted and guidelines developed to create video material for HRI video trials. It was beneficial

---

[1] Even in scenarios like ours where the robot's behaviour does not depend on the subject's responses the live situation *affords* interaction.

that both first person and third person views were used for the video trials as we think that this enabled subjects to get a more realistic perspective of space and distance for the robot.

The current findings offer scope for future work into the feasibility of using video based HRI trials to aid the design and implementation of live interaction studies. We have only considered one human-robot interaction scenario in the current study. It is important to determine whether these results can be replicated and in addition to consider different scenarios. In some ways, the current trial was challenging, as the issues of speed, space and distance were considered. It may be the case that video footage is more comparable to live trials for exploring subject responses to robot gestures, robot appearance, and robot dialogue.

In addition to replicating the current study, a range of different set-ups could be considered. In the meantime, we carried out studies into subjects' opinions towards robot approach directions when the subject is standing rather than sitting, and whether they are sitting behind a desk or not when the robot approaches them with an object. We also used video footage in HRI studies that investigated opinions toward different robotic appearances (e.g. mechanistic vs. humanoid appearance). As the research area of socially interactive robots is relatively new, there are few design paradigms to provide input to this project and others [18]. However, until more studies are carried out using different scenario set-ups and robot behaviours with larger sample sizes, we cannot yet positively conclude that video based HRI trials are a reliable and informative means of gaining data to assist in the future design of robot companions.

IV. CONCLUSIONS

To conclude, encouraging results were obtained comparing the agreement between subject responses towards robot approach directions for live and video human-robot interactions. This has positive implications for researchers designing future HRI trials, as video trials could be used as a complementary research tool to yield valuable results regarding peoples' opinions towards various aspects of a robot's behaviour and/or physical capabilities. Video trials are more economical compared to live interactions, and allow the designers/researchers greater levels of control and standardisation over the set-up of the trials, which is sometimes difficult when conducting live HRI trials. We hope that our results, while still at a preliminary stage, will open up discussions on the design space of HRI experiments and methodologies.

V. ACKKNOWLEDGEMENT

Many thanks go to our colleagues for participating in our trials; in particular, to Mike Blow for his acting and narration skills, and also his patience, in helping to produce the HRI video scenarios.

REFERENCES

[1] Fong, T., Nourbakhsh, I. and Dautenhahn, K. A survey of socially interactive robots. *Robotics and Autonomous Systems*, *42*. 143-166.

[2] Te Boekhorst, R., Walters, M.L., Koay, K.L., Dautenhahn, K. and Nehaniv, C., A study of a single robot interacting with groups of children in a rotation game scenario. in Proc. IEEE *CIRA 2005*, (Finland, 2005). 35-40.

[3] Walters, M.L., Dautenhahn, K., Te Boekhorst, R., Koay, K.L., Kaouri, C., Woods, S., Nehaniv, C., Lee, D. and Werry, I., The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment. in *Proc. IEEE RO-MAN 2005,* (Nashville, USA, 2005), 347-352.

[4] Walters, M.L., Dautenhahn, K., Te Boekhorst, R., Koay, K.L., Kaouri, C., Woods, S., Nehaniv, C., Lee, D. and Werry, I., The influence of subjects' personality traits on predicting comfortable human-robot approach distances. in *Proc. COG SCI 2005; Towards Social Mechanisms of Android Science Workshop*, (Stresa, Italy, 2005), 29-37.

[5] Severinson-Eklundh, K., Green, A. and Hüttenrauch, H. Social and collaborative aspects of interaction with a service robot. *Robotics and Autonomous Systems*, *42*. 223-234.

[6] Kanda, T., Hirano, T., Eaton, D. and Ishiguro, H. Interactive robots as social partners and peer tutors for children: A field trial. *Human Computer Interaction, 19* (1-2). 61-24.

[7] Robins, B., Dautenhahn, K. and Dubowski, J., Investigating Autistic children's attitudes towards strangers with the theatrical robot - A new experimental paradigm in human-robot interaction studies. in *Proc. IEEE RO-MAN 2004*, (Kurashiki, Japan, 2004), 557-562.

[8] Walters, M., Woods, S., Koay, K.L. and Dautenhahn, K., Practical and methodological challenges in designing and conducting interaction studies with human subjects. in *Proc AISB'05*, (University of Hertfordshire, Hatfield, UK, 2005), 110-119.

[9] Hinds, P., Roberts, T. and Jones, L. Whose job is it anyway? A study of human-robot interaction in a collaborative task. *Human Computer Interaction*, *19*. 151-181.

[10] Dario, P., Guglieimelli, E. and Laschi, C. Humanoids and personal robots: Design and experiments. *Journal of Robotic Systems*, *18* (12). 673-690.

[11] Kanda, T. and Ishiguro, H. Reading human relationships from their interaction with an interactive humanoid robot. *Lecture Notes in Computer Science*, *3029*. 402-412.

[12] Kanda, T., Ishiguro, H., Imai, M. and Ono, T., Body movement analysis of human-robot interaction. in *IJCAI 2003*, (2003), 177-182.

[13] Kidd, C.D. Sociable robots: The role of presence and task in human-robot interaction *Massachusetts Institute of Technology*, Massachusetts, 2003.

[14] Shinozawa, K., Naya, F., Yamato, J. and Kogure, K. Differences in effect of robot and screen agent recommendations on human decision-making. *IJHCS*, *62* (2). 267-279.

[15] Paiva, A., Dias, J., Sobral, D., Aylett, R., Woods, S., Hall, L. and Zoll, C., Caring for Agents that care: Building empathic relations with synthetic agents. in Proc. *AAMAS Autonomous Agents and Multi Agent Systems*, (Columbia University, New York City, USA, 2004)

[16] .Murray, L. and Trevarthen, C. Emotional regulation of interaction between two-month-olds and their mothers. in Field, T. and Fox, N. eds. *Social perception in infants*, Norwood, NJ, USA, 1985, 101-125.

[17] Nadel, J., Carchon, I., Kervella, C. and Marcelli, D. Expectancies for social contingency in 2-month-olds. *Developmental Science*, *2* (2). 164-173.

[18] Dautenhahn, K., Woods, S., Kaouri, C., Walters, M., Koay, K.L. and Werry, I., What is a robot companion - Friend, assistant or butler? in *Proc. IEEE IROS*, (Edmonton, Canada, 2005), 1488-1493.

Naturally Occurring Gestures in a Human-Robot Teaching Scenario: An exploratory study

Nuno Otero, Chrystopher Nehaniv, Kerstin Dautenhahn,

Joe Saunders, Aris Alissandrakis

School of Computer Science, Adaptive Systems Research Group, University of
Hertfordshire, College Lane, Hatfield, Hertfordshire, AL10 9AB, UK

December 2005

1 Introduction

This report describes our initial steps investigating how naturally occurring human gestures can be used to facilitate the interaction and communication between humans and robots.

This work is being developed as part of the research for the European funded project COGNIRON. More specifically an exploratory user study was performed to illuminate which naturally occurring gestures can be observed in a scenario specifically relevant to experimental tasks relevant for the project. The term ‘naturally’ here refers to an unconstrained scenario where subjects were not given any scripts or pre-defined gestures to use. Within the COGNIRON project, the line of research described in this report conforms to the objectives defined for Work Package 3.4 - “Requirements for contextual interpretation of body postures and human activities”. The main focus of this report, in order to capture these requirements, is the on-going development of a coding scheme to classify gestures people produce when asked to demonstrate how to perform a task. The coding scheme is an essential part of our strategy to systematically study the frequency, duration and sequence of different gestures in people's task demonstrations. In turn, this analysis will inform the design and parameterisation of algorithms for the recognition of human activities. This work was inspired by previous research in COGNIRON, specifically: Nehaniv et al. (2005) provided the conceptual framework for the coding scheme categories and research by University of Karlsruhe (2004) informed our present work in relation to requisites for human activity descriptions from a system's perspective.

The line of work is being pursued in close partnership with University of Karlsruhe. However, we hope the annotated corpus will be available for inspection by the scientific community as a tool for informing research in this topic.

Considering the context of the COGNIRON project, we are not focusing on the recognition of expressive displays or the *production* of expressive behaviour, such as facial and emotional expressions. Nevertheless, we do acknowledge the importance of emotions in the modulation of communication events and interactions between humans and computer artefacts.

The rest of this report will go as follows. First, we will summarise relevant research concerning the role of gestures in the interaction process with a particular emphasis on how the results from human-human interaction informed (or can inform) the development of computational artefacts. Secondly, we will describe the exploratory study run. We then proceed with the presentation of the results. The third part consists of a general discussion and topics for future research.

2 Background

Human-Robot Interaction (HRI) is a recent research field. Kiesler and Hinds (2004) consider that the study of design alternatives to facilitate Human-Robot Interaction is a new focus of Human-Computer Interaction (HCI), although Dautenhahn (2004), Robins et al. (2005) and Nehaniv et al (2005) argue that HRI will need to develop its own discipline specific methods due to the embodied nature of interaction with robots. Kiesler and Hinds (2004) consider that although many insights and methodologies developed in HCI might be applicable in HRI, robots pose specific design issues:

- Some research suggests that humans tend to anthropomorphize robots more than other computational artefacts.
- The fact that robots are mobile and will interact with humans in a common physical space highlights the need to understand how to model the sharing and "negotiation" of space. In other words, robots will have to negotiate their interactions in a dynamic (and we would add social) environment.
- Robots will have to make decisions and these decisions might have an immediate impact at the physical and/or social level.

The theme of HRI has been growing in importance as robots are moving out of the confines of industrial settings. In fact, several technological breakthroughs are paving the way for an expansion of robots' use in different contexts and performing diverse tasks, sometimes in close collaboration with humans (Kiesler & Hinds, 2004).

Dautenhahn (1998) points out that the idea of agents being able to interact with humans in a "natural" way is considered attractive. This statement seems especially relevant when discussing the concept of believability. As robots start acting in human environments issues of agency, believability and sociality become very important. In fact, it seems that increasing one of these factors (using some kind of design strategy) influences the others. Robots that inhabit human social spaces will need to be designed to conform as much as possible to human expectations. The interactions need to be "acceptable" and "comfortable" to humans (Dautenhahn, 1998). Fong, Nourbakhsh and Dautenhahn (2003) state that the design of sociable robots needs input from research concerning social learning and imitation, gesture and natural language communication, emotion and recognition of interaction patterns. Moreover, according to these authors, three primary types of dialogue are crucial to foster the robots' abilities to interact with humans: low level interaction mechanisms, non-verbal communication, and natural language.

To some extent, our present research focus, as stated in the introduction, can be described as an investigation of the specificities of gestures for interacting with robots. For our starting point the questions are: How do humans use gestures in their communication processes? How should we proceed to find design solutions for robots that take advantage of these human abilities? Let us turn our attention to the first question.

Gestures are closely linked with the accompanying speech in terms of timing, meaning and communicative function (see, for example, Cassell, 2000; Kendon, 1997; McNeill, 1992). Adam Kendon's and David McNeill's work on the study of human communicative gestures are considered to be a landmarks in the field (see Kendon, 1997; see McNeill, 1992, for an overview). McNeill (1992) takes a very restricted definition of gestures as he considers them to be only the non-manipulative hand/arm movements that occur during speech. Kendon (1997), however, considers difficult to specify what kinds of body movements (in a broad sense) should be called gestures. Nevertheless, he presents boundaries for what he considers to be gestures: "...only actions that are treated by co-participants interaction as part of what a person meant to say will be included: conventional gestures, gesticulations, and signing are included, but posture shifts, self-touchings, and incidental object manipulations are not." (Kendon 1997, pag. 110). Kendon's proposal (which in fact follows from his earlier work in the 80s) defines a broader set than the strict definition put forward by McNeill. In fact, McNeill (1992) refers to Kendon's definition as the *Kendon's continuum*: Gesticulation → Language-like Gestures → Pantomimes → Emblems → Sign Languages. According to McNeill (1992):

> *"As we move from left to right: (1) the obligatory presence of speech declines, (2) the presence of language properties increases, and (3) idiosyncratic gestures are replaced by socially regulated signs." (pag. 37).*

Gestures can be classified into distinct types and different classificatory systems do exist[1]. For example, McNeill (1992) identifies the following types:

- Iconics are gestures that convey a close formal relationship to the semantic content of speech. These gestures transmit information regarding some property of the speech referent.

- Metaphoric gestures are similar to iconic ones. However, in this type of gesture the hand/arm movement presents and abstract ides and illustrate the speech content using a third element that acts as a metaphor.

- Beats - these gestures tend to have the same form regardless of the content. They resemble the beating of musical time. Furthermore, this type of gestures is characterised by having two phases while other types have three: preparation, stroke and retraction. Beats can index the level of significance of words in the flow of speech. They are valuable for their discourse-pragmatic content.

- Deictics are basically pointing gestures that indicate objects, persons and events in the real world, as well as abstract concepts to control the flow of a conversation.

---

[1] A review of the different classification systems is beyond the scope of this work. For a starting point into the topic see, for example, McNeill (1992); Kendon (1997), Cassell (2000); and Kipp (2004).

- Cohesives are gestures that show logical connections between different parts of the discourse.
- Emblems are gestures with culturally established meaning. These can be used on their own unambiguously, that is, independent of speech. However, they are frequently culturally bound.

Cassell (2000) adds one type: propositional gestures. People use these gestures to convey information that can be said to be true or false, one example: "I caught a fish this big []" where the gesture (denoted by []) intends to show the actual size of the fish.

Kipp (2004), while specifying the different types of gestures considered being relevant for the development of an embodied conversational agent, includes one more type: adaptors. Adaptors are non-communicative self and object touches (examples: scratching ones head, fiddling with a pen). According to Kip (2004) some researchers do not consider these to be gestures due to the non-communicative and intentional nature. They claim that the relation to speech is very loose. However, Kipp (2004) considers that adaptors have informational value since they give indications concerning the speaker's state. Furthermore, adaptors can also segment or give clues about the flow of the communication act.

McNeill (1992) considers gestures to be global-synthetic. With these terms the author highlights the relationship between parts and whole. In gestures, the meaning is only extractable if related to the whole. Furthermore, McNeill (1992) also states that gestures do not combine in a hierarchical fashion. However, Kipp (2004) considers that, although not frequent, there are examples that contradict this assertion.

The study of gestures' function has been diverse. Research in problem solving has found not only that gestures can convey specific information and reveal thoughts not revealed in speech but also that observing gestures can be a useful extra to speech when trying to uncover cognitive processes (Alibali, Bassok, Solomon, Syc, & Goldin-Meadow, 1999; Garber & Goldin-Meadow, 2002). However, in fact, research investigating the function of gestures in relation to speech has produced contrary results or divergent opinions. For example, some authors defend the importance of gestures semantics independently from speech (Cassell, McNeill, & McCullough, 1999; Kendon, 1997) while others consider that gestures primary function is not to convey semantic information (Krauss, Dushay, Chen, & Rauscher, 1995). Gestures have also been studied in relation to child development and language acquisition (Iverson & Goldin-Meadow, 2005; Ozcaliskan & Goldin-Meadow, 2005), or teaching and learning strategies (Kelly, Singer, Hicks, & Goldin-Meadow, 2002; Roth & Lawless, 2002; Singer & Goldin-Meadow, 2005).

Let us now turn to the second question posed above: How should we proceed to find design solutions for robots that take advantage of these human abilities?

The use of gestures in the communication loop between humans and computer artefacts has been explored. A simplified overview of research on multimodal interfaces and human-robot interaction seems to indicate the following approaches for the inclusion of gestures in the interaction process:

- Some research has explored the use of sets of pre-defined gestures and speech to communicate with robots (Ghidary, Nakata, Saito, Hattori, & Takamori, 2002; Oh, Lee, & You, 2005; Severinson-Eklundh, Green, & Huttenrauch, 2003).
- Other researchers consider that the use of gestures for communication with computer artefacts can and should be explored beyond the confines of a set of pre-defined gestures (Cassell, 1998, 2000; Cassell & Thorisson, 1999; Kipp, 2004; Koop, Tepper, Ferriman, & Cassell, in press; Nehaniv et al., 2005).

Robots may need to recognize human gestures and movements, infer limited intent regarding these human gestures and movements and to communicate their own internal state and plans to humans (Nehaniv et al., 2005). In their paper, the authors analyse the specificities of analysing and incorporating gestures in the interaction loop between humans and robots and argue for the need to consider a broad definition of gestures. The reason being to avoid:

> *"...inherent limitations of approaches working with a too narrow notion of gesture, excluding entire classes of human gesture that should eventually be*

> *accessible to interactive robots able to function well in a human social environment." (Nehaniv et al. (2005), pag. 372).*

Therefore we need to adopt a broader notion of gesture than the one proposed by McNeill (1992) or even Kendon (1997). To illustrate the point made, take the example of manipulative gestures. These gestures involve the manipulation of objects (usually without concomitant speech) and seem central to HRI since a robot, at times, will have to co-ordinate object manipulations with humans. Should we not include hand/arm movements that exchange objects between partners involved in interaction as a type of gesture? For purposes of human-robotic interaction, we shall adopt the following operational definition:

A gesture is any bodily kinematic transition effected by an agent (human or robot). These may be interactional, communicative (or not), symbolic (or not), or manipulative, etc. without prior restriction.

We assume kinematic transition of the body is effected by the agent, i.e. self-initiated unless there is evidence to the contrary. Thus gesture under this definition includes gesticulation when speaking, manipulation, eye gaze transitions, and in principle even such kinematic transitions as are involved in blinking, or in speech. Treating speech and "communicative gestures" together is in line with current neuroscience views involving the role of mirror neurons in the evolution of human language (Rizzolatti & Arbib, 1998).

In practice, focus will usually be limited to visible gestures that can be detected with given sensors, and acoustical gesture will be treated generally under the category 'speech'. Furthermore, generally, the focus of the notion gesture will be restricted upper body and torso effectors (with speech treated separately, e.g. as a parallel track).

This notion of gesture is in line with Nehaniv et al. (2005), who propose the following five classes of gestures:

- Irrelevant or Manipulative gestures - these are gestures do not have a primary communicative or interactive function (in practice, this class is split).

- Side Effect of Expressive Behaviour - these are gestures that occur as side-effects of people communicative behaviour. It can be motion with hands, arms, face etc but without specific interactive, communicative, symbolic or referential roles.

- Symbolic gestures - these are gestures that follow a conventionalised signal. Its recognition is highly dependent on the context (both current task and cultural milieu).

- Interactional Gestures - this category classifies gestures used to regulate interaction with a partner. Thus are can be used to initiate, maintain, invite, synchronise, organise or terminate an interaction behaviour between agents.

- Referencing/pointing gestures - the gestures that fall into this category are gestures used to indicate objects or loci of interest.

Nehaniv et al. (2005) stress the importance of knowing the context in which gestures are produced since this is crucial to disambiguate meaning. The authors add that data on the interaction history and context help the classification process. The authors also point out the need to consider to whom or what is the gesture targeted (identify target) and who, if anyone, is supposed to see it (identify the recipients). Certain gestures in particular situations might be multipurpose (Nehaniv et al., 2005). For example, a gesture of bringing an object conspicuously and not overtly toward an interaction partner is manipulative but it may also be classified as interactional since it might be solicitation for the partner to take the object. This issue resonates with what McNeill (1992) termed as the global-synthetic property of gestures - the attribution of meaning is only extractable if related to the whole (see above).

As referred to in the introduction, for the present study we followed the classification system proposed by Nehaniv et al. (2005) for the elaboration of a coding scheme to identify people's gestures when asked to explain a home task to a robot. The next two sections describe the study.

3 The exploratory study

This initial and exploratory study had two goals: a) we wanted to start developing and testing a coding scheme to accurately classify the types of gestures people produce when asked to demonstrate how to perform tasks, and b) we also aimed at finding out to what extent robots' appearance influenced the way people use different gestures to explain a certain task to a robot. The robot's appearance was divided into three categories: (a) humanoid robots, (b) animal-like robots, and (c) mechanistic robots.

The experiment had two phases. In the first one, participants were asked to sort a series of robots' photographs into distinct groups and label them - *open card sorting task*. The idea was to assess to what extent our initial categories of the robots' appearance were indeed relevant for the participants and give us some insight on the relevance of some appearance features of each robot.

In the second phase, participants were asked to instruct a robot on how to perform a certain task using only gestures and also to actually demonstrate how it should be completed. The task involved: (a) taking some plates from a cupboard, (b) setting the plates and corresponding cutlery as if one were setting a table and finally (c) pick up the plates again and put them away. The level of granularity for breaking the overall task into sub-tasks for the explanation and corresponding gesturing and demonstration was at object manipulation level, since the participants were instructed to manipulate only one object at a time. This means that the participants had to show precisely how each object was manipulated in accordance to the overall goal of the task. They were asked to perform in front of video camera and imagine that this was in fact the vision system of the robot. The participants were prompted to consider the different types of robots by being shown the photographs previously utilised in the card sorting task, but this time in groups sorted beforehand by the experimenters.

3.1 Methodology

This study followed a within-subjects design: the participants had to demonstrate how to perform the task for the three different types of robots under consideration - the experimental conditions. Considering that there were three conditions, the order of appearance was counterbalanced to try to cancel effects of order. Thus we had 6 arrangements of the three conditions[2].

3.1.1 Participants

The sample consisted of 12 participants: 7 men and 5 women, all from our research institute, either researchers or staff (this choice of subjects seemed appropriate due to the preliminary nature of this first trial). Six of the participants belong to our research group but the other six are either researchers from other groups or members of the administrative office.

3.1.2 Apparatus / Materials

Cards with photos of different robots

A set of 18 cards with photographs of different robots were given to the participants. These photographs showed six exemplars of the three categories of the different types of robots' appearance we referred to in the beginning of this document (see Appendix 7.3). So, we acknowledge possible bias concerning the choice of photographs shown.

Questions for the semi-structured interviews

The following questions were asked at the end of the session as part of a semi-structured interview:

- What do you think about this experiment? Did you find it interesting?
- How did you feel about it? Did you enjoy or not? Did you feel engaged with the task or bored?

---

[2] We should note, however, that although this design tried to control the effect of order it assumes that the possible effects of the order are identical in all the three conditions. For example, it assumes that transiting from human-like to animal-like will have the same effect as transiting from animal-like to mechanistic look, and so forth. In order to test such assumptions we would need a bigger sample.

- Which kind robot do you think you would like most? Would you try to buy one if the cost was not prohibitive?
- Can you realistically see yourself doing something similar in real life?
- In relation to the task of teaching the robot, do you think you would like to take time to learn some kind of gestures for communicating with the robot?
- Do you think the robot should be able to learn the gestures you would like to use to communicate with it? In case you are not the only person who interacts with the robot, would you spend time defining the gestures with other people?
- Do you think you were influenced by the different robots' appearances? If yes, in what way?

Software for giving feedback to participants

A software program was created to simulate the robot's feedback to participants. The feedback was simulating the robot's understanding of the gestures produced by the participants. The feedback consisted on the display of three colours in a computer screen. The computer screen image was projected with a video projector to a wall in front of the participants (see sub-section 3.1.5 for a description of the physical setting).

The three colours and corresponding meanings were: a) *red* if the system did not understand at all the meaning of the gestures produced; b) *yellow* if the system understood partially but further specification was needed; and c) *green* if the gestures were understood. However, the actual display of the feedback followed a random generation of sequence of colours with a probability of: 20% for red, 20% for yellow and 60% for green. One of the experimenters was controlling the segmentation of when to display feedback by pressing a button at the end of each sequence of the participants' gestures.

Coding scheme for the analysis of participants video recordings

Initial exploratory studies are essential for the development of coding schemes (Bakeman & Gottman, 1997; Robson, 2002). Kipp (2004) considers four different types of way for the transcription of gestures:

- Structural transcription involves the identification of the gestures phases. It seems to be common agreement that gestures can be decomposed into three phases: preparation, stoke and retraction. Kipp (2004) discusses in detail the phases and the implications for gesture classification.

- Descriptive transcription involves the detailed description of the gesture's form. With this approach the gesture's form is encoded through the specification of the position joints at a point in time. This procedure is carried out for regular time intervals. This approach needs no subjective judgements from the observer/coder.

- Functional transcription entails ascribing meaning to the gesture being displayed. According to Kipp (2004) "Meaning and function can be inferred by carefully observing the gestures and their accompanying speech" (pag. 53). It is an interpretative approach.

- Categorial transcription - in this case the gestures are classified according to a set of categories. The categories are described in detail and exemplars are given to facilitate the identification.

The different transcription approaches are not exclusive. In fact, in a coding scheme more than one can co-exist.

Descriptional tracking of sensory data or joint angles is often available to roboticists as raw un-interpreted description of human gesture and activity. However, we are concerned with requirements for robotic systems for interpreting these data in a meaningful and useful way for Human-Robot Interaction. In our case our coding scheme is a mix between functional and categorial transcriptions.

The first version of the coding scheme for the video recording analysis was based on previous conceptual work by Nehaniv et al. (2005) and further developed after careful observation of the video data collected (see Appendix 7.1, for a description of the categories used). At this stage we did not use rules to disambiguate the classification. The observers/coders had the description of the categories and attributed the categories to the behaviours according to their interpretation. However, the observers/coders did watch the video of one of the participants together and discussed the classification as a way to train their coding skills and agreement.

Additional materials

The following material was also used:

- a video camera - to record the participants gestures,
- a sound recorder - to record the interviews at the end of the session,
- a laptop - to control the feedback program,
- a video projector - to show the feedback to the participants,
- a set of plates and cutlery.

3.1.3 Procedure

3.1.3.1 Description of the physical setting

The room utilized had enough table space to accommodate the participants' demonstrations. The actual physical layout is presented in Figure 1.

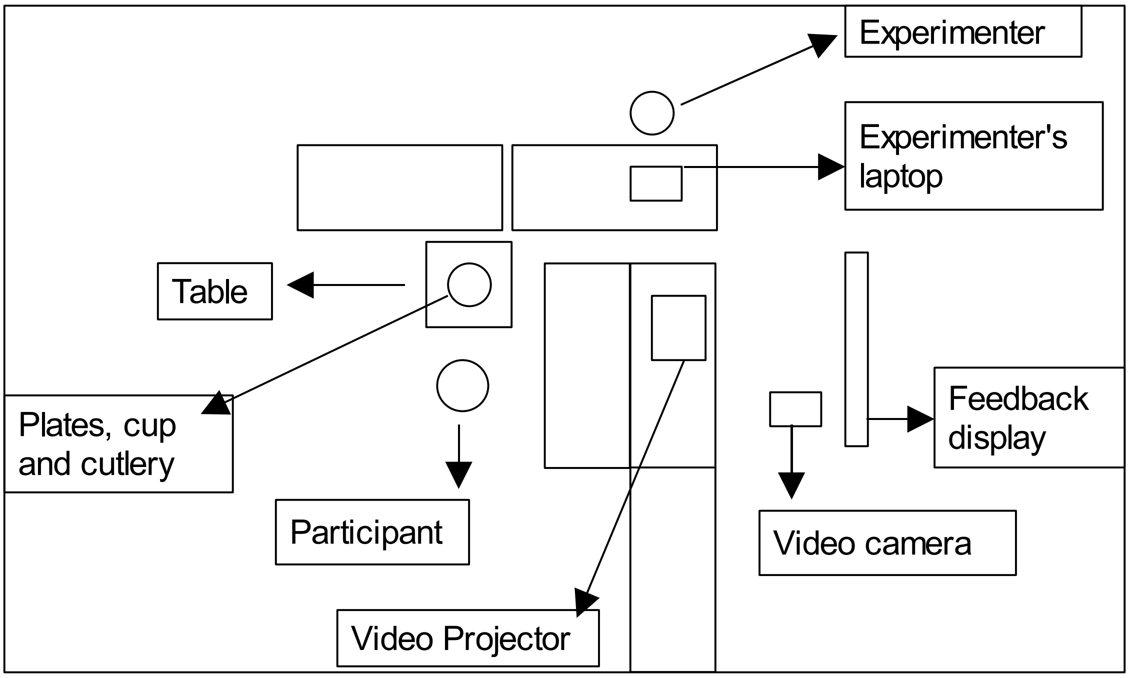


Figure 1 - Layout of the experiment room

3.1.3.2 Spoken instructions to participants and sequencing of activities

Step 1 - beginning of the experiment

- INSTRUCTION - "This experiment has three parts. In the first part you will be asked to sort a set of photographs of robots into different groups you find relevant. In the second part, you will be asked to explain to three different robots how to accomplish a certain task. In the final part, we will ask you a few questions concerning your experience with our experiment."

- The nature of the different materials in the room was explained to the participant.

Step 2 - The initial card sorting task.

- INSTRUCTION - "Please look at the following photos of robots and try to sort them into groups that you find meaningful. There is no right or wrong answer to this task."

- After the completion of the task asked the participants the rationale for the sorting.

Step 3 - The main experimental task.

- INSTRUCTION - "Our experiment involves recognising gestures to control a robot. This camera is the robot's vision system. It will recognise gestures and tell you if it has recognised them successfully. Your performance is not being assessed. Our robot's vision is being assessed. Now consider that this is the type of robot you will have to teach". The experimenter re-presents the cards of the type of robot to be taught and leaves them on the table. "The way to do it is:
    - Suppose my colleague wanted to gesture to the robot to move the chair.
    - First the robot says: WAITING…….
    - You make the gesture.
    - The robot responds with a colour.
    - If RED, the robot has not understood at all and will wait – please try again
    - IF YELLOW, the robot has partially recognised the gesture and will wait – please try again
    - If GREEN, the robot HAS understood. - You must now actually move the object.
    - Thus WAIT, GESTURE, get COLOUR, TRY AGAIN or MOVE OBJECT."
- INSTRUCTION - "Only one object at a time can be manipulated. There are three tasks. The first one is: move the plates to the table; you may place the plates in any position on the table. The second one is: Lay second table using the plates and the cutlery. The third one is: put plates and the cutlery back onto the stand."

- The procedure was repeated for the other two types of robots.

4 Results

4.1 The card sorting task

Eleven of our participants came up with 25 different designations for the groups of photographs they formed. From this initial number of designations we were able to aggregate into the groups being displayed in Table 1.

Considering the four main categories (humanoid, mechanistic, animal, toys), the participants seemed to agree fairly well on which robots' photographs to include in the groups they formed. However, two photographs revealed to be more difficult to classify: the photograph showing robot Care-O-bot (by IPA) was classified either as humanoid or mechanistic, while the photograph showing the robot EMIEW (by Hitachi, Ltd) was classified as toy or humanoid.

Participants' Categories	Our aggregation
Human	Humanoid
Realistic Human	Humanoid
Abstract Human	Humanoid
Mechanistic	Mechanistic
Tool	Mechanistic
Precise	Mechanistic
Lab Exp	Mechanistic
Prototypes	Mechanistic
Experimental	Mechanistic
Industrial	Mechanistic
Assembly Line	Mechanistic
Manufacturing	Mechanistic
Animal	Animal
Biological	Animal
Dogs	Animal
Insects	Animal
Toys	Toys
Cute	Toys
Joky	Toys
Pods	No classification
Mystery	No classification
Commercial	No classification
Animal/Human	No classification
Have Legs	No classification
Have Faces	No classification

Table 1 - Participants' designations for the groups of photographs formed and our initial aggregation

4.2 Analysing the videos

The videos were analysed by two observers/coders using the coding scheme referred to previously. From the initial sample of 12 participants, only 9 were picked for the analysis of their video recordings. The observers/coders utilized the software Annotate to code the activity of the participants. This tool requires the coder to choose a time interval (granularity) for the coding. In this case the time interval we considered practical was one second. As we shall see, however, this choice had clear implications on the coding. Annotate assigns one letter corresponding to the particular event type in the category chosen by the observer/coder to the time interval. More than one letter can be assigned to the same time interval since the tool supports non-mutually exclusive category systems. Annotate produces a table (in a text file format) for each subject with columns for each category and rows for the time intervals.

This section, the analysis of the video recordings, is sub-divided into four parts, each part corresponding to the measure utilised. The measures were:

- Counts – this simply means the frequencies of the different types of behaviours.
- Duration – this means duration of each type of behaviour. The grouping was made by duration interval and not type of behaviour (so we have, frequencies of behaviours that lasted 1, 2, 3, 4, 5, 6, 7, 8 seconds). In this case we are not considering sequences of types of behaviours.
- Multiple occurrences – whenever more than one code appears at a specific moment (row of the spreadsheet) then we register a multiple occurrence. Given that some behaviours seem to happen very fast and our time interval for coding is 1 second then it might be the case that sometimes we are registering multiple occurrences when we should be registering sequences.
- Repetitions – in this case we analyse if the type of sequence that happens when the participants are asked to repeat their gestured explanation is different from what they produced initially. It basically tries to see to what extent the participants

try different types of sequences or if they stick to a particular type. For the analysis we considered changes involving duration or type of sequence. As examples of the heuristic followed for the classification: (a) if the initial sequence was ABBA and the repeated sequence was AABBAA then it was classified as a change in duration, (b) if the initial sequence was ABB but the repeated sequence shows BAA then we would classify it was change in type. Please note that due to the way Annotate writes the output file of the coding activity, it is almost impossible to distinguish between duration and multiple consecutive occurrences of equal events. In other words, AAA might be a type A behaviour lasting three seconds or three consecutive occurrences of A. In order to disambiguate the issue we would need a special code to distinguish it.

- Sequences – the sequences were classified based on the type of behaviours appearing from the onset to the offset of the sequence but not considering the duration of each type of behaviour. This means that: ABBBA was considered similar to ABA but different from AABAB. The same applies here regarding the confounding between duration and multiple consecutive occurences referred to in repetitions.

4.2.1 Frequency of the different behaviours

The registries of referencing object and referencing place are inflated for E, L, R and S. This reflects the strategy of one of the coders to consider that eye gaze almost always meant referencing (to the objects, to the display, to places). The other coder did not follow the same strategy. This suggests rather than interpreting eye gaze as either referencing or looking, it might be more useful to simply track the target of eye gaze in the annotation.

Let us start by considering the results of participants A, Af, B, D, K, all coded by the same observer/coder (please consider the paragraph above, regarding one of the observer/coder strategy; this issue is further discussed in the Discussion and conclusions section). The results from Table 2 suggest that participant A used more referencing object and referencing place. The other participants used more miming manipulation and miming transportation.

	Participants								
Categories	A	Af	B	D	E(*)	K	L(*)	R(*)	S(*)
Ref. Obj.	**78-49.7%**				**80-28.8%**		**99-26.9%**	**121-30.1%**	**119-45.9%**
Ref. Plc.	**77-48%**	**7-3.1%**			**72-25.9%**		**120-32.6%**	**123-30.6%**	**123-47.5%**
Mm Trsp		**77-34.2%**	**73-46.8%**	**85-48.9%**	**63-22.7%**	**76-47.8%**	**74-20.1%**	**78-19.4%**	**2-0.8%**
Mm Mnp		**76-33.8%**	**74-47.2%**	**85-48.9%**	**62-22.3%**	**77-48.4%**	**72-20%**	**19-19.7%**	**15-5.8%**
Symb Gst		**58-25%**							
Gst Crt		**4-1.4%**							
Exprss	**1-0.6%**		**4-2.5%**				**1-0.3%**	**1-0.2%**	
Incid.	**1-0.6%**	**3-1.3%**	**5-3.2%**	**2-1.1%**	**1-0.4%**	**6-3.8%**	**2-0.5%**		

Table 2 - Absolute frequencies and percentages for each type of behaviour taking into consideration the relevant categories for the analysis of the gesturing phase. The categories' defintions can be found in appendix 7.1
Legend - Ref Obj: referencing object; Ref plc: referencing place; Mm Trsp: miming transportation; Mm Mnp: miming manipulation; Symb Gst: symbolic gesture; Gst Crt: gesture creation; Exprss: expressive gesture; Incid: incidental gesture
***Note - the interactions of participants A, Af, B, D and K were coded by one observer while participants E, L, R, S were coded by another observer. The registries of referencing object and referencing place are inflated for E, L, R and S. Please see explanation below in the main text. This reflects the strategy of one of the coders to consider that eye gaze almost always meant referencing (to the objects, to the display, to places). The other coder did not follow the same strategy.**

This suggests subjects used two main different strategies: use of referencing or use of miming. Table 2 also shows that only one subject used symbolic gestures extensively. We did not register many occurrences of unintentional or expressive gestures.

We believe that the same applies to the other participants coded by the other observer/coder. The confounding factor is the way he registered the referencing behaviour. In fact, participant S is also a good example of the use of referencing object and place since the number of miming behaviour is rather small.

4.2.2 Duration of the different behaviours

The estimates regarding duration are constrained by the time interval chosen for the coding. In fact, in some cases, we could observe that the participants' gestures took less than one second. So, regard the following results as approximations and not precise accounts.

Figure 2 suggests that referencing objects seems to take about one second. However, this estimate might be confounded due to the way one of the observers/coders utilised this category (looking at the display and looking at the plate, cup etc). Referencing place also seems to take one second most of the times (see Figure 3).

Miming grasping and transportation is mostly a one or two second gestures but there were a few cases of three second gestures (see Figures 6 and 7). The same applies to the actual manipulation and transportation of the objects (see Figures 4 and 5). Finally, the subject who used symbolic gestures seemed to take between 1 to 3 seconds to do them.

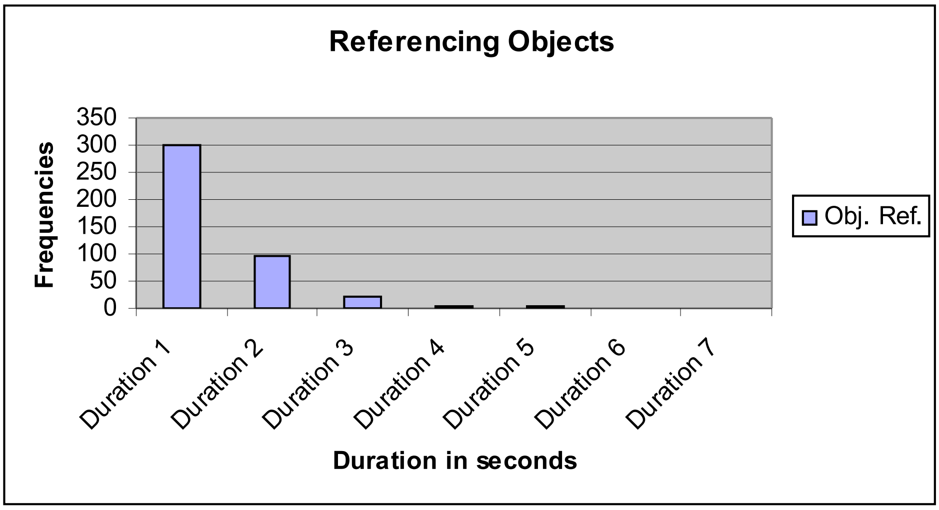


Figure 2 - Duration of the referencing object behaviour

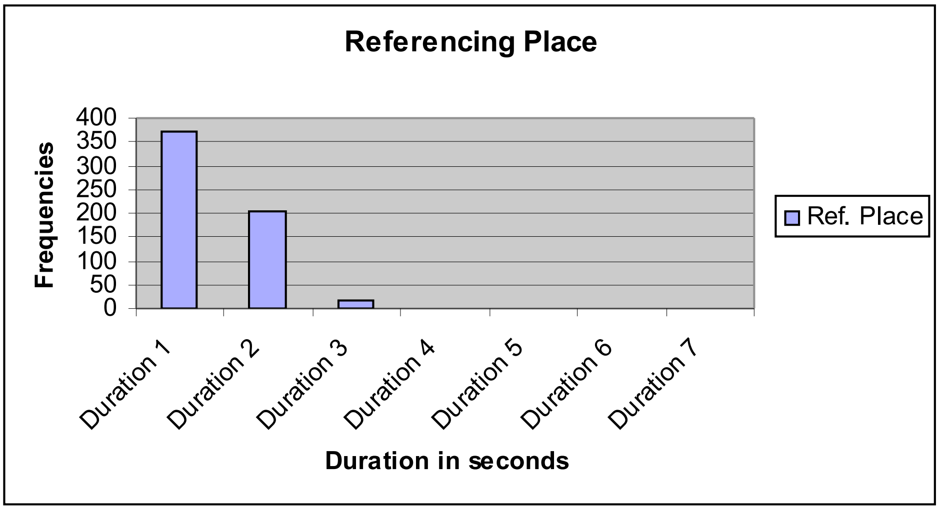


Figure 3 - Duration of the referencing place behaviour

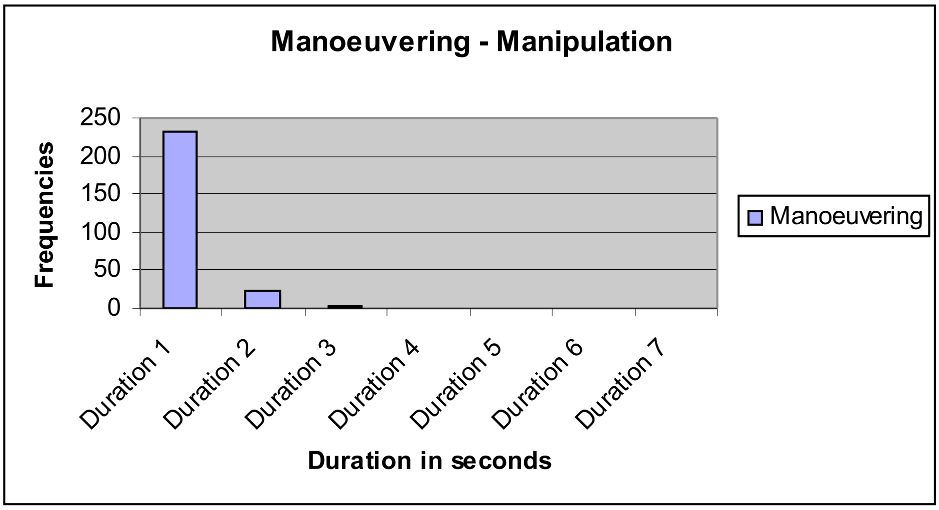


Figure 4 - Duration of the actual grasping of the objects

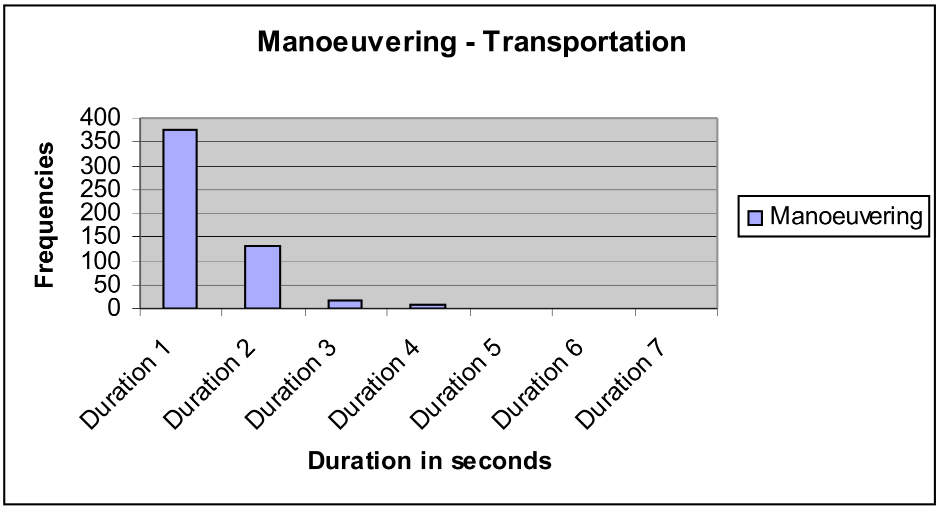


Figure 5 - Duration of the actual transportation of the objects

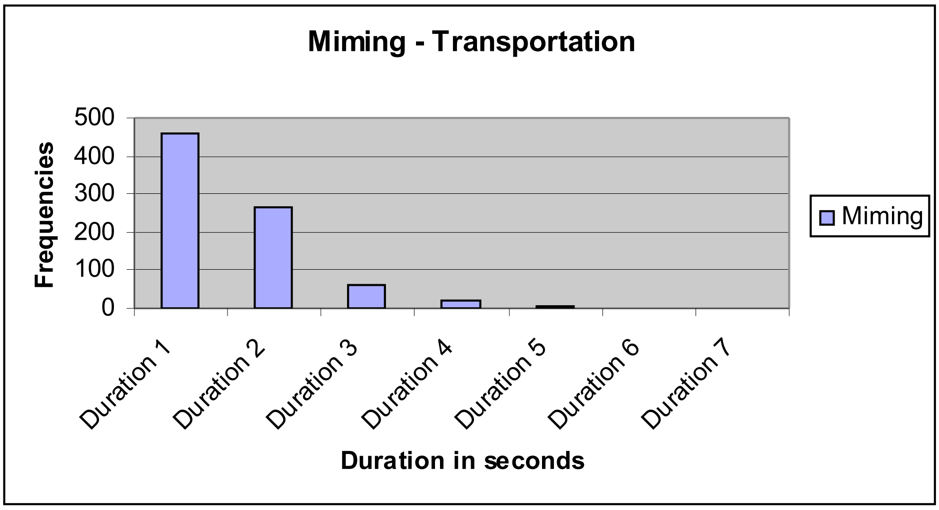


Figure 6 - Duration of the miming of object transportation

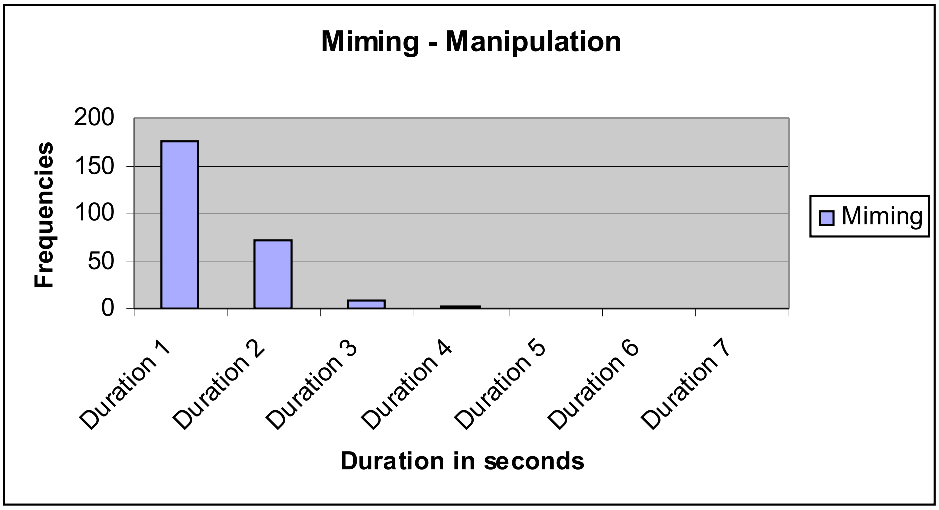


Figure 7 - Duration of the miming of object grasping

4.2.3 Multiple occurrences

Given the way one of the observers/coders coded referencing it is difficult to distinguish between referencing objects and the meaning of eye gaze. The observer/coder that did not use the same strategy only registered 12 multiple occurrences, 11 of which were from the same subject using object referencing at the same time as transportation.

4.2.4 Repetitions

Figure 8 shows that the subjects seem to stick with the same strategy along the session since whenever they are asked to repeat their demonstration they tend only to do it slower. Only two subjects, E and S, have more changes to the type of sequence than to the duration. This result reinforces the finding concerning the use of two primary strategies: use of referencing or use of miming.

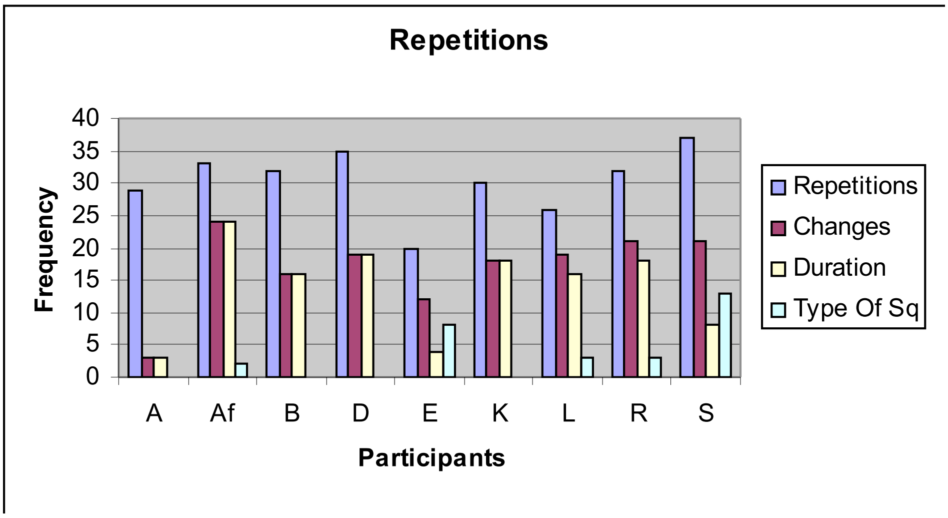


Figure 8 - Absolute frequencies of repetitions, changes to the previous sequence and type of change

4.2.5 Sequences of behaviours

Considering that the two observers/coders had different strategies regarding the coding of referencing, the analysis of sequences is separated.

	Participants				
Sequences	A	Af	B	D	K
Manipulation+Transportation		21	71	85	76
Gst Crt+Manipulation+Transportation		4			
Symbol+Manipulation+Transportation		52			
Manipulation+Transportation+Referencing Place		6			
Manipulation+Transportation+Express			2		
Referencing Object+Place	74				

Table 3 - Frequencies of the different type of sequences

According to one of the observers/coders coding (participants A, Af, B, D, and K) the main findings are (see Table 3):

- Participant A used just referencing. The sequence was referencing object first and place after.
- Participants B, D and K used almost exclusively miming.
- Participant Af not only mixed miming with referencing but also displayed the greater variety of distinct sequences.

Taking into consideration the way the other observer/coder coded referencing (we should bear in mind the implications of his strategy in the registering of multiple occurrences) we chose to create four additional categories for the analysis of data so coded. The new categories are the result of combining referencing with miming as co-occurring (see Table 4).

	Manipulation	Transportation
Referencing		
Object	Type 1(T1)	Type 3(T3)
Place	Type 2(T2)	Type 4(T4)

Table 4 - the four categories resulting from the combination of miming and referencing

The main findings are (see Table 5 and Figure 9):

- The most frequent sequence was referencing object and manipulation (T1) followed by referencing place and transportation (T4).
- Participant E almost equally displayed two different types of sequences: (a) the previously described sequence and (b) referencing place and manipulation (T2) followed by referencing place and transportation (T4). This participant also displayed the highest variety of different sequences.
- Participant S, who used referencing did it mostly in the following order: referencing object followed by referencing place.

	Participants			
Sequences	E	L	R	S
T1+T3	8		1	
T1+T4	17	48	72	
T2+T3	4			
T2+T4	21	8		
T3+T4	1			
T1+T3+T4	4	9	2	
T1+T2+T4	4	8		
T1+T3+T2	1			
T2+T3+T4	2			
T1+T4+T2				1
T2+T1+T4				1
Ref Obj+Ref Place				60
Ref Plc+Ref Objt				4
Ref Obj+Ref Place+ Ref Obj				1
Ref Place+Ref Obj+Ref Place				4

Table 5 - Frequencies of the different types of sequences produced by the participants

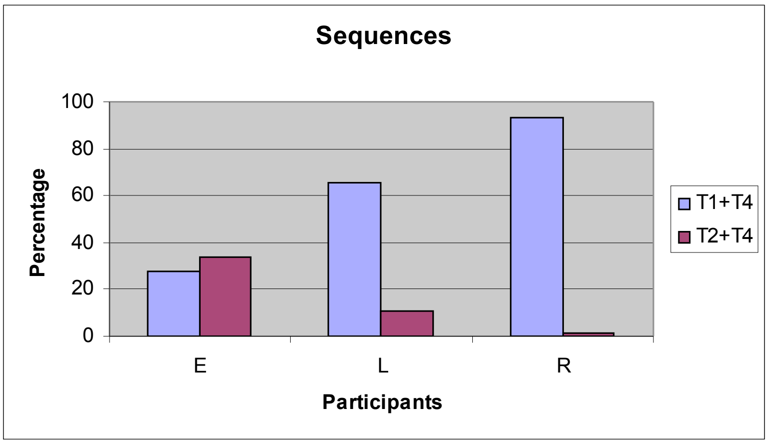


Figure 9 - Percentage of the two most frequent sequences in relation to the overall number of sequences each participant produced

4.3 Analysing the interviews

Question 1 and 2 - What do you think about this experiment? Did you find it interesting? How did you feel about it? Did you enjoy or not? Did you feel engaged with the task or bored?

For the analysis of the participants' views concerning questions 1 and 2 we decided to consider five categories of contributions: the task was boring, repetitive, interesting, engaging and other. For each participant's whole answer more than one category can be used.

In relation to the distinction between "interesting" and "engaging", we coded engaging whenever the participants explicitly stated the word or whenever they provided expressions that meant they were actively exploring possibilities or ideas while doing it.

From Table 6 we can see that some of the participants judged the task repetitive. In fact, three of them labelled it boring. However, we could find 6 answers judging the task interesting and 3 of them engaging. It was particularly interesting to find out that one of participants clearly stated of being actively and methodically searching for correlations between its gesturing and the feedback.

Categories	Freq.	Highlights
Boring	3	"A little bit too long"
Repetitive	4	
Interesting	6	"It was not boring. I liked doing it" "It was interesting, not boring... just repetitive"
Engaging	3	"It is different to actually think about a task you do naturally"; "I liked doing it. I tried various combinations. I tried logical orders." "It raised lots of questions while I was doing it".
Other	2	"I didn't have feedback about the robot's performance." "I wasn't quite sure about what was meant by gestures."

Table 6 - Frequencies of the categories chosen and highlights of the participants' answers

Question 3 - Which kind robot do you think you would like most? Would you try to buy one if the cost was not prohibitive?

The analysis of this question's answers is simple: we just register the participants' preferences. It is worth noting, however, that the majority of the participants seem to have adopted our categorization, or at least something similar. It is worth noting that no participants choose the mechanistic type.

Categories	Freq.	Highlights
Humanoid	4	"In particular because of the tasks involved." "I preferred the one it was expressive and a bit funny."
Mechanistic		
Animal	3	"They look cute."
"Cute" or Toy	2	"The ones that look less technological."
"Designed"	1	"Not mechanical"

Table 7 - Frequencies of the categories chosen and highlights of the participants' answers

Question 4 - Can you realistically see yourself doing something similar in real life?

In relation to question four, the main point we chose to consider was: to what extent do the participants see themselves performing a teaching task to a robot similar to the one presented to them.

	Freq.	Highlights
Yes, I could do it	8	"The only thing I was slightly confused was whether or not the gestures I was producing were appropriate". "... much more with the humanoid..."
No, I couldn't	2	"Not to lay a table. I would do it myself."
Other	3	"If I was buying a robot I would expect that someone would have trained it." "I would be probably doing the task and explaining what I was doing" "I would expect to be able to recognise this simple task (...). Not using gestures"

Table 8 - Frequencies of the categories chosen and highlights of the participants' answers

Question 5 and 6 - In relation to the task of teaching the robot, do you think you would like to take time to learn some kind of gestures for communicating with the robot? Do you think the robot should be able to learn the gestures you would like to use to communicate with it? In case you are not the only person who interacts with the robot, would you spend time defining the gestures with other people?

The analysis of question five focuses on whether the participants prefer the use of a pre-defined set of gestures or teaching the robot their own. Furthermore, we also wanted to see to what extend people were aware of the implications of having a robot at home. In particular, to what extent they thought they would have to create a common vocabulary to communicate with the robot.

Use manual	Freq.	Highlights
Yes	10	"If it was an easy manual... not just read-read-read." "Yes, because if I had a robot I would try to get the best of it."
No	0	
Other	2	"Maybe, it depends how many pages or how practical. If the robot was very efficient learning..." "I think it is simpler to do it by gesturing."

Table 9 - Frequencies of the thought use (or not) of a manual with pre-defined gestures and highlights of the participants' answers

Table 9 shows that most of the participants would be willing to read a manual and learn about a set of pre-defined gestures. However, on Table 10 we can see that the sample is split between people that would prefer to use a set of pre-defined gestures and people that would go for a "teaching gestures" approach. It seems that people that chose the pre-defined gestures were concerned with the time needed to teach the robot. People that would prefer to teach the robot seem to argue that this system seems more natural.

Preference	Freq.	Highlights
Manual	4	"If it has a specific set of gestures that it was trained to recognise it should be easier. The other way I might get frustrated thinking there is an easy way to do this." "I would prefer to have a set of gestures pre-defined." "It depends on what tasks..."
Teach/ Gestures	6	"I think showing gestures is a bit easier "I would prefer this sort of system (the gesturing). It is close to how people teach somebody else."
Other Method	1	"Maybe a two way thing with the manual" - gestures and manual.

Table 10 - Frequencies of the preferred method to demonstrate the task to a robot and highlights of the participants' answers

In relation to the issue of co-ordinating the set of gestures to teach the robot if more than one person was involved, only four participants clearly addressed the theme. All considered that they would be able to co-ordinate with other people. However, two of them also stressed that they think the robot should be able to be flexible enough to accommodate to different people. For example, one of the participants said: "I think the robot should be flexible. With simple sign gestures the essence is always the same. It should be able to recognise that different people were doing the same gesturing. It was a bit surprising to find that the majority of the participants did not discuss the issue more fully. Two possibilities come to mind: (a) the question was not properly understood; (b) the participants considered that their gestures were self explanatory for that task so they could not see the point of the need to find common ground.

Question 7 - Do you think you were influenced by the different robots' appearance? If yes, in what way?

In this last question we considered the following analysis:

- Do the participants report being influenced by the robots' appearance?
- Do the participants report thinking about the robots' appearance?
- What types of robots fostered distinction?

Influenced?	**Freq.**	**Highlights**
Yes	6	"Yes, the mechanical ones influenced. I thought I had to be more mechanical, making less smooth gestures." "There was this thing I thought half way through, especially when looking at the spiders."
No	3	"It is a matter of internal design." "Although I tried to differentiate and look at the pictures I did not get influenced. I thought of a brain and the outer case wouldn't make a difference."
Thought about it	1	"I couldn't work out how to make my gestures different."

Table 11 - How many times did the participants report being influenced by the robots' appearance?

Type of robot	**Freq.**	**Highlights**
Mechanistic	1	"I thought I had to be more mechanical, making less smooth gestures."
Humanoid	0	
Animal	5	"There was this thing I thought half way through, especially when looking at the spiders." "I couldn't figure out how the biologically inspired robots could reproduce my gestures." "I was influenced by the animal type. I was curious about how they could reach and hold objects etc." "For toy robots or insects I would have to demonstrate. But for the humanoid I should be able to just point."

Table 12 - What type of robots' appearance the participants report to influence their actions or thinking?

5 Discussion and conclusions

The results from the card sorting task suggest that there are some kind of underlying grouping principles regarding the robots' features that people use to classify robots since people agree on photographs' groups to be formed given the set provided. What are these features, however, seems to be open to discussion. To our knowledge there is no systematic study addressing this topic.

In relation to the coding scheme developed, the task set up for the second phase of the experiment was simple and it was expected that the recognition of the gestures' meaning should not be difficult. However, the observers/coders had some trouble coding the gestures. We are aware that the two observers/coders seemed to have different strategies towards the coding scheme that influenced the coding. For example, one of the coders considered that eye gaze almost always meant referencing (to the objects, to the display, to places) while the other did not.

These two examples clearly indicated the need to improve the description of the different categories in order to raise inter-observer/coder agreement. In support of the tentative nature of this first study, some research points out the difficulties subjects have of unambiguously interpreting gestures alone (Hadar & Pinchas-Zamir, 2004).

In terms of lessons learned, we would like to point out the need to clearly disambiguate between duration *versus* multiple and consecutive occurrences of the same behaviour. One way to solve this is the use of an event-based annotation: it explicitly shows the onset and offset of events based on the available frames of data. For our next codification tasks we plan to use an alternative video coding software tool that is event-based - ANVIL (Kipp, 2004). Furthermore, ANVIL, being an event-based video annotation tool, does not require the *a priori* specification of a particular time interval.

Another issue is the "usability" of the classification system itself, from the perspective of the coders. For coding purposes it seems easier to have nested categories since it facilitates the search for codes (see Kipp, 2004, for a discussion of annotation tool requirements). The ANVIL tool is flexible in relation to the use of a multi-level coding scheme and allows to creation of relationships between the levels and/or layers utilised for the coding. ANVIL generates an xml file as output with all the coding information. This seems extremely useful to proceed with the analysis of the data both quantitatively and qualitatively.

A second version of the coding scheme is being defined and will be tested regarding inter-observer agreement. In fact, we are planning to extend our work and produce a fully standard coding scheme with corresponding manual This second version differs from the previous one on the following aspects (for the description of the coding scheme and its different categories, see Appendix 7.2):

- We will explicitly require the definition of the context and tasks being carried out.

- We will explicitly separate in different coding tracks movements with hand/arm from facial displays, head movements and verbalisations. This means that the classification system now has more categories.

- We will consider linking some of the tracks corresponding to different categories in order to foster the extraction of meaning.

From the analysis of video recordings the following results are particularly suggestive and need further research:

- We were unable to register, using the coding scheme, differences in people's gestures due to exposure to the distinct groups of photographs. This is particularly evident since people tended to stick to the same type of sequence along the whole session. Nevertheless, it was interesting to note that in the interviews some of the participants report being influenced or having thought about it. Some participants, however, explicitly stated they were not being influenced by the different photographs. We feel further research would be needed to clarify this topic. Maybe using photographs is not enough? Maybe if we have actual robots the results would be clearer or at least the two positions (being or not being influenced) can be made more distinct. Another possibility is that our present coding scheme was not detailed enough to capture the possible subtleties. For example, a brighter picture might emerge if we classify the movements people make according to a dimension of smoothness/jerkiness[3]...

- For this particular task two styles seem to emerge: either people used referencing object and then place or they just mimed the task.

- The duration of most gestures falls in an interval of one to three seconds. It is interesting to observe, though, that miming behaviours seem to take slightly longer than the actual manipulation of the objects. The question is: if the

---

[3] The gestures some of the participants produced suggested to observers some association of jerky movements to mechanistic-like robots and smooth movements to human-like robots.

participants were not told to gesture first and move objects after would the same relation hold?

- People do not seem to switch their style of gesturing not even when negative feedback regarding its understanding was given. Instead they just slowed down a bit their explanatory sequence. However, one question comes to mind: would this be the case if instead of having a robot the recipient was another human? Is it possible that people were assuming that the non-understanding from the vision system was just a perceptual problem? Another possibility is that people just assumed that such a simple task would entail an "obvious" sequence and were not really creative due to the nature of the task requirement.

From the interviews we highlight:

- People see themselves teaching a robot using gestures.
- People seem to accept using a manual to learn a set of pre-defined signs to be able to communicate with the robot.
- Preference towards either using a manual to learn how to communicate with the robot *versus* teaching the robot using gestures is almost split in half.
- People report being influenced by the different types of robots' photographs, in particular when considering the animal-like robots.

Our initial study focused on the production of gestures alone. On one hand, it seems reasonable to consider that people may not find it “natural” to use gestures alone for the explanation of tasks to a robot. On the other hand, research points out that people's explanations using gestures alone of assembling tasks seems to facilitate not only their comprehension of the task but also the comprehension of the observers (recipients) of their explanations (Lozano & Tversky, 2004, 2005). Hence, further research is needed to understand the interactions of roles gesture and speech might assume in human-robot interaction.

For our next study we intend to set up an experiment to investigate the use of speech and gestures when explaining simple tasks related to Key Experiments in Cogniron. The experiment will follow a within subjects design where each subject will be asked to explain the task using gestures and speech or gestures alone. The participants will also be asked to answer a questionnaire regarding their experience using gestures and speech and how they picture this kind of interaction happening between humans and robots. The experiment will allows us to see to what extent speech and gesture combine in this particular situation and collect data regarding the participants' opinions on this particular style of interaction.

To wrap up, we argue that this exploratory study was essential for the on-going development of a coding scheme to classify gestures people produce when asked to demonstrate how to perform a task. The coding scheme is central for the systematic study the frequency, duration and sequence of different gestures in people's task demonstrations. In turn, this analysis will inform the design and parameterisation of algorithms for the recognition of human activities.

6 References

Alibali, M. W., Bassok, M., Solomon, K. O., Syc, S. E., & Goldin-Meadow, S. (1999). Illuminating mental representations through speech and gesture. *Psychological Science, 10*(4), 327-333.

Bakeman, R., & Gottman, J. M. (1997). *Observing Interaction: An Introduction to Sequential Analysis*. Cambridge: Cambridge University Press.

Cassell, J. (1998). A Framework For Gesture Generation And Interpretation. In R. Cipolla & A. Pentland (Eds.), *Computer Vision in Human-Machine Interaction* (pp. 191-215). New York: Cambridge University Press.

Cassell, J. (2000). Nudge Nudge Wink Wink: Elements of Face-to-Face Conversation for Embodied Conversational Agents. In J. Cassell & J. Sullivan & S. Prevost & E. Churchill (Eds.), *Embodied Conversational Agents* (pp. 1-27). Cambridge, Massachusetts: The MIT Press.

Cassell, J., McNeill, D., & McCullough, K. E. (1999). Speech-Gesture Mismatches: Evidence for One Underlying Representation of Linguistic and Non-Linguistic Information. *Pragmatics and Cognition, 7*(1), 1-33.

Cassell, J., & Thorisson, K. R. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence, 13*(4-5), 519-538.

Dautenhahn, K. (1998). The Art of Designing Socially Intelligent Agents: Science, Fiction, and the Human in the Loop. *Applied Artificial Intelligence, 12*, 573-617.

Dautenhahn, K. (2004). Robots we Like to Live With? - A Developmental Perspective on a Personalized, Life-Long Robot Companion, *Proceedings Of The IEEE Ro-man, 13th International workshop on Robot and Human Interactive Communication* (pp. 17-22). Kurashiki, Okayama, Japan: IEEE Press.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems, 42*(3-4), 143.

Garber, P., & Goldin-Meadow, S. (2002). Gesture offers insight into problem-solving in adults and children. *Cognitive Science, 26*(6), 817-831.

Ghidary, S., Nakata, Y., Saito, H., Hattori, M., & Takamori, T. (2002). Multi-Modal Interaction of Human and Home Robot in the Context of Room Map Generation. *Autonomous Robots, 13*(2), 169-184.

Hadar, U., & Pinchas-Zamir, L. (2004). The semantic specificity of gesture - Implications for gesture classification and function. *Journal Of Language And Social Psychology, 23*(2), 204-214.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science, 16*(5), 367-371.

Karlsruhe, U. o. (2004). *Report on a classification of human activities in households environments (D.2.3.1).*

Kelly, S. D., Singer, M., Hicks, J., & Goldin-Meadow, S. (2002). A helping hand in assessing children's knowledge: Instructing adults to attend to gesture. *Cognition And Instruction, 20*(1), 1-26.

Kendon, A. (1997). Gesture. *Annual Review Of Anthropology, 26*, 109-128.

Kiesler, S., & Hinds, P. (2004). Introduction to This Special Issue on Human-Robot Interaction. *Human-Computer Interaction, 19*(1&2), 1-8.

Kipp, M. (2004). *Gesture Generation by Imitation: From Human Behaviour to Computer Character Animation*. Boca Raton, Florida: Dissertation.com.

Koop, S., Tepper, P., Ferriman, K., & Cassell, J. (in press). Trading Spaces: How Humans and Humanoids use Speech and Gesture to Give Directions. *Spatial Cognition and Computation*.

Krauss, R. M., Dushay, R. A., Chen, Y., & Rauscher, F. (1995). The Communicative Value of Conversational Hand Gestures. *Journal of Experimental Social Psychology, 31*, 533-552.

Lozano, S., & Tversky, B. (2004). Communicative Gestures Benefit Communicators. In K. Forbus & D. Gentner & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 849-854). New Jersey: Lawrence Erlbaum Associates.

Lozano, S., & Tversky, B. (2005). Gestures Convey Semantic Content for Self and Others. In B. Bara & L. Barsalou & M. Bucciarelli (Eds.), *Proceedings of the Twenty Seventh Annual Conference of the Cognitive Science Society* (pp. 1337-1342). New Jersey: Lawrence Erlbaum Associates.

McNeill, D. (1992). *Hand and Mind.* Chicago: Chicago University Press.

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

Oh, J. Y., Lee, C. W., & You, B. J. (2005). Gesture recognition by attention control method for intelligent humanoid robot, *Knowledge-Based Intelligent Information And Engineering Systems, Pt 1, Proceedings* (Vol. 3681, pp. 1139-1145).

Ozcaliskan, S., & Goldin-Meadow, S. (2005). Do parents lead their children by the hand? *Journal Of Child Language, 32*(3), 481-505.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Cognitive Sciences, 21*, 188-194.

Robins, B., Dautenhahn, K., Nehaniv, C., Mirza, N. A., Francois, D., & Olsson, L. (2005). Sustaining interaction dynamics and engagement in dyadic child-robot interaction kinesics: Lessons learnt from an exploratory study, *Proc. IEEE Ro-man 2005* (pp. 716-722).

Robson, C. (2002). *Real World Research*. Oxford: Blackwell Publishing.

Roth, W. M., & Lawless, D. (2002). Scientific investigations, metaphorical gestures, and the emergence of abstract scientific concepts. *Learning And Instruction, 12*(3), 285-304.

Severinson-Eklundh, K., Green, A., & Huttenrauch, H. (2003). Social and collaborative aspects of interaction with a service robot. *Robotics and Autonomous Systems, 42*(3-4), 223.

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science, 16*(2), 85-89.

7 Appendix

7.1 The coding scheme

For the building of the present set of categories, three coders watched the videos of three participants, discussed an initial categorization system, defined the coding scheme and extended the categorization system whenever the coding scheme revealed imprecise. The categorisation system follows the Nehaniv's et al (2005) work.

At the end of this document you can find a table with the set of categories to be used for the analysis utilising the Annotate software.

7.1.1 The categories and coding scheme

Reference/topicalization

The gestures that fall into this category are gestures used to indicate objects or loci of interest. We chose to clearly set these apart from gestures that involve the demonstration of object manoeuvring. This means that if a person shows a grasping action this will be classified as manoeuvring and manipulation and not some type of referencing.

The category reference/topicalization is further subdivided into to sub-categories and corresponding elements:

- **Object of reference/topicalization:**
    - **Objects - when the reference is made to an object.**
    - **Place - when the reference is made to a place (physical space).**
    - **Self - when the reference is made to one-self.**
    - **Other persons - when the reference is made to other persons.**
- **Body part used for the referencing:**

    - **Head/eyes**
    - **Arm/hand/fingers**

Manoeuvring gestures

The classification of manoeuvring gestures is applied when the gesture corresponds to the actual body movement utilised for the manipulation of artefacts. In other words, these gestures imply some kind of effect over artefacts and can be performed not also with the hands (for example, grasping an object) but also with other body parts (for example, kicking a ball).

The category of manoeuvring gestures can be sub-divided into two distinct subcategories whether the artefact is actually being manipulated or the person just demonstrates how to do it. Thus the sub-categories are:

- **Actual manoeuvring**
    - **Transportation - when involves the displacement of the artefact.**
    - **Manipulation - for example, tool use, grasping, pulling a lever.**
- **Miming**
    - **Transportation - same as above.**
    - **Manipulation - same as above.**

Body/World relationship

This category describes the state of the person's body in relation to the world. It can be sub-divided into:

- **Still**
    - **Sitting**

    - **Standing**
- **Moving**
    - **Walking**
    - **Running**
    - **Other...**

Symbolic gestures

These are gestures that follow a conventionalised signal. Its recognition is highly dependent on the context (both current task and cultural milieu). We chose to sub-divide this category into:

- **Symbolisation of objects**
- **Symbolisation of actions**
- **Symbolisation of abstract concepts**
- **Symbolisation of person or animal**
- **Symbolisation of an adjective**

Furthermore, we can also further classify the gestures according to the following categories:

- Arbitrary
- Iconic
- Indexical
- Metaphorical

Interaction regulation

According to Nehaniv et al (2005) this category classifies gestures used to regulate interaction with a partner. Thus are can be used to initiate, maintain, invite, synchronise, organise or terminate an interaction behaviour between agents. For our present purposes we chose not only to sub-divide this category according to the focus of the gesture but to classify according to its function. The sub-categories are:

- **Focus (it involves not only gaze direction but also the use of other body parts, for example using the head)**
    - **Object**
    - **Place**
    - **Person or animal**
- **Function**
    - **Initiating**
    - **Awaiting acknowledgment**
    - **Maintain**
    - **Invite**
    - **Synchronise**
    - **Organise**
    - **Terminate**

Gestural symbol creation

This category will be utilised whenever the coder observes the person performing a gesture that is in fact the creation of an un-conventionalised signal.

Expressive gestures

According to Nehaniv's et al (2005) categories expressive gestures occur as side-effects of people communicative behaviour. It can be motion with hands, arms, face etc but without specific interactive, communicative, symbolic or referential roles.

Unintentional/irrelevant activities

Unintentional/irrelevant gestures do not have a communicative or interactive function. They might prompt an observer to make inferences about the internal state of the other. For example, "he is not paying attention" but the transmitter is unaware of its effect on the observer. For example, scratching one's head, playing with objects.

Repetition

This category just covers the effect of giving feedback to the person given the experimental task set ut.

Error and error recovery

Finally, the error and recovery category intends to code actions that the observer can clearly judge to be errors or recovery from previous errors in relation to the on-going task.

7.1.2 Annotate table

Categories and sub-categories			**Annotate code**
Reference/Topicalization			
	Object of Reference		**Obj. Ref.**
		Object	*A*
		Place	*B*
		Self	*C*
		Others	*D*
	Body part used		**Bdy Part**
		Head/eyes	*A*
		Arm/hand/finger	*B*
Manoeuvring gestures			
	Actual		**Man Actl**
		Transportation	*A*
		Manipulation	*B*
	Miming		**Man Dmns**
		Transportation	*A*
		Manipulation	*B*
Body/World Relation			
	Still		**Bd/Wrld/Sti**
		Standing	*A*
		Sitting	*B*
	Moving		**Bd/Wrld/Mv**
		Walking	*A*
		Running	*B*
		Other	*C*
Symbolic Gestures			**Symb. Gest**
	For objects		*A*
	For action		*B*
	For abs. concept		*C*
	For person/animal		*D*
	Adjective		*E*
Interaction Regulation			
	Focus		**Inter Focus**
		Object	*A*
		Person	*B*
		Place	*C*
	Function		**Inter Fnc**
		Initiating	*A*
		Maintain	*B*
		Invite	*C*
		Synchronise	*D*
		Organise	*E*
		Terminate	*F*
		Awaiting Acknowledgement	*G*
Gesture Creation			**Gest Crt - A**
Expressive gestures			**Exprssv - A**
Unintentional/Irrelevant			**Unint - A**
Repetition			**Repete - A**
Error/error recovery			**Error - A**

Table 13 - table of the categories and corresponding codes for the Annotation software
Note - Giving the need to flatten the sub-categorization into two levels for using Annotate, the code in bold correspond to the categories used and the italic letters A, B, C, D etc are the sub-categories

7.2 Work in progress: The new coding scheme

This coding scheme intends to be generic; it should be possible to adapt to different study requirements. Researchers will be able to choose tracks, and categories adapting to the required level of detail.

A Context

A1 Who is present?

The initial description of who (or what) is present (humans, robot, distinguishable groups of agents) allows the identification of the recipient.

A2 Objects present

Allows the identification of the target, in terms of targets for manipulation, topics of discourse etc...

A3 Layout

Describes the physical setting and helps contextualise the movements and manipulations performed.

A4 Scenario/Script

This includes description of the tasks and typical activities.

A5 Interaction history

This includes information or pointers to previous episodes involving some agents or objects, as well as recent activities.

B Track for Gestures - (using hands or other body parts except head)

B1 Irrelevant gestures

Irrelevant gestures that do not have a communicative or interactive function. They might prompt an observer to make inferences about the internal state of the other. For example, "he is not paying attention" but the transmitter is unaware of its effect on the observer. For example, scratching one's head. This category also includes side effects of motor activity, e.g. rhythmic movement of arms while walking.

B2 Manipulative gestures

These gestures imply some kind of effect over artifacts and can be performed not also with the hands (for example, grasping an object) but also with other body parts (for example, kicking a ball). The coder will need to identify the actual **manipulandum**, and if appropriate the **recipient** and the **target** and/or **target locus**.

Consider whether to decompose into **phases** or not - for example, preparation, grasp and transportation / displacement.

B3 Symbolic gestures

Might designate

- **objects** – state which object (maybe referring to A2 but not exclusively);
- **actions** – state which action. The coder will need to identify for actions taking argument roles, the **recipient** and the **target** (or other roles).
- **person / animal / other agent** - identify referent .
- **other** such as abstract concepts, uses of non-referential symbol (e.g. adverbials for ‘hurry up’, ‘no’, refusal, greeting, numbers, etc.) - state interpretation.
- If it is the first occurrence of a gesture observed for that person, not culturally prescribed as far as the coder is aware, then he should code **gesture creation**.

B4 Referential / pointing gestures

State which place, object and/or agent(s) (persons, humans or groups) (maybe referring to A1, A2 and A3 but not exclusively). These gestures might also point to parts of the body that are themselves displaying a gesture. For example, in one hand I start by symbolising an abstract concept and then, with the other hand, I point to it. So, it seems reasonable to assume that people can also refer to absent, abstract or imaginary things.

B5 Interactional gestures

Function to the interaction episode: initiating, awaiting acknowledgement, maintain, synchronize and terminate.

This category might be used in conjunction with some of the others described in this document - multipurpose gesture. In particular, miming, manipulation, referential, symbolic and/or head movements.

B6 Side effect of expressive behaviours

According to Nehaniv's et al (2005) categories expressive gestures occur as side-effects of people communicative behaviour. It can be motion with hands, arms, face etc but without specific interactive, communicative, symbolic or referential roles.

B7 Miming

State the **action** but also which **object**(s) and/or agent that might be in the role of argument(s) as for manipulative actions (maybe referring to A2 but not exclusively).

Consider whether to decompose into **phases** or not - for example, preparation, grasp and transportation / displacement.

C Track for Facial Displays and Head Movements

State if the classified event is congruent to other co-occurring events: ***congruent / non-congruent****. For example, eye gaze is referring to the same object as pointing.*

C1 Emotional Displays

Possibly annotate also - Function to the interaction episode: initiating, awaiting acknowledgement, maintain, synchronize and terminate.

C2 Eye gaze (identify target of gaze - object, interaction partner, location, etc)

Possibly annotate also - Function to the interaction episode: topicalize (target of gaze), initiating, awaiting acknowledgement, maintain, synchronize and terminate.

C3 Head movements

For example, head nods, “saying” no etc...

D Track for Verbal Productions

State if the classified event is congruent to other co-occurring events: ***congruent / non-congruent****. For example, naming an object and pointing to it at the same time.*

D1 Sentences uttered.

Possibly annotate also - Function to the interaction episode: initiating, awaiting acknowledgement, maintain, synchronize and terminate.

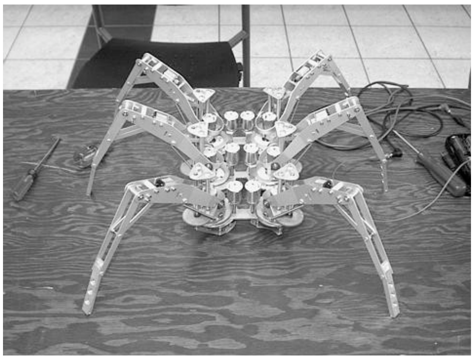
E Track for Body / World relationship

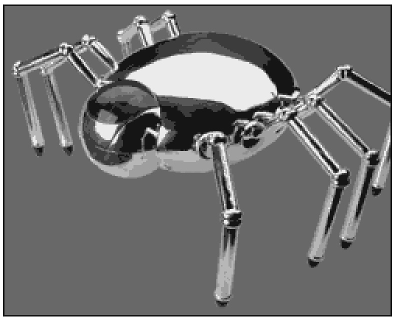
This category specifies the position of the body in the physical space as well as changes:

- General posture:
    - Seated

    - Standing
- Movement - state changes of location:
    - Walking
    - Running
    - Smooth or jerky

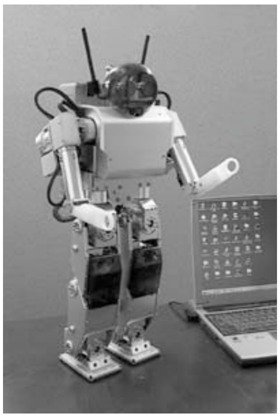
7.3 The robots' photographs

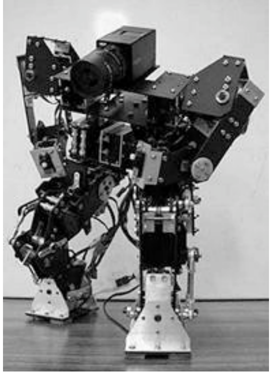


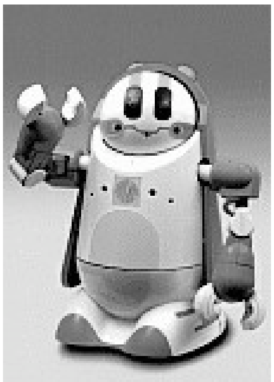


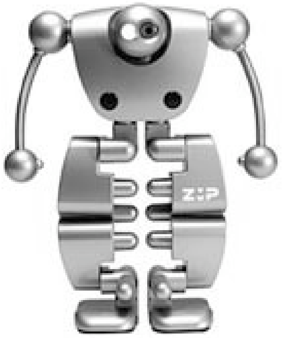

PeopleBot

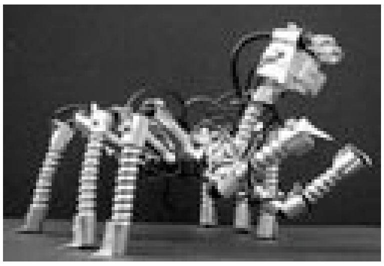

(c)Hitachi, Ltd. 2005. All rights reserved.











Investigating Spatial Relationships In Human-Robot Interaction

Helge Hüttenrauch, Kerstin Severinson Eklundh, Anders Green, Elin Anna Topp*
Royal Institute of Technology (KTH) *Centre for Autonomous Systems
School of Computer Science and Communication
Stockholm, Sweden
+46 790 91 56

{hehu, kse, green, topp}@nada.kth.se

ABSTRACT

Co-presence and embodied interaction are two fundamental characteristics of the command and control situation for service robots. This paper presents a study of spatial distances and orientation of a robot with respect to a human user in an experimental setting. Relevant concepts of spatiality from social interaction studies are introduced and related to Human-Robot Interaction (HRI). A Wizard-of-Oz study confirms that the preferred spatial distance and spatial formations encountered can be modeled to this framework, but equally reveal novel patterns of spatial cooperation in *interaction episodes*. Furthermore, it is claimed that a simplistic parameterization and measurement of spatial interaction miss the dynamic character and might be counterproductive in the design of socially appropriate robots.

Keywords

Human-Robot Interaction, spatiality in interaction, interaction episodes

1. INTRODUCTION

Humans engaged in physical activities have to deal with spatial relationships. Due to the physical mass and degrees of freedoms of body, head, and limbs, movement in three-dimensional space or manipulation of objects needs to be orchestrated on the basis of sensory perception and cognitive abilities. The required understanding of spatiality is claimed [20] to have its origin in evolutionary traits that shaped not only perception, but influenced human usage of linguistic metaphors in daily usage, too.

A service robot that is intended to operate in co-presence of a human user constitutes a physical entity on its own. Consequently do both, i.e. the human interactor and the robot need to deal with issues stemming from their mutual co-presence, mobility, multimodal communication, and embodied interaction [7].

Trained by sociocultural and daily experience, humans are in general skilled in dealing with other people in managing space or in handling objects. The signaling, whether through non-verbal or verbal expressions is often well understood (or can be repaired ad hoc).

Figure 1: User teaching the robot objects, illustration only

As a requirement for this ability it is mandatory for people to notice other people's body movements; gaze exchanges, gestures or mimic expressions are also closely tracked and interpreted subconsciously. Furthermore signaling in and through the environment is possible and anchored within the sociocultural context and practice, e.g. a closed door can signify a "please do not disturb me right now" if this convention is well established and adhered to.

Interactive mobile robots are machines that challenge or defy many innate human assumptions: They are pushing the borderline of our understanding, differentiation, and reactions towards what is *alive* or *inanimate* [18], [23]. Their self-locomotion, the attribution of "body"-movement as expression of their own intention, and the missing training in reacting to a machine's spatial behavior are contributing factors.

As part of the European research project COGNIRON [5], this paper investigates the spatial management in a Human-Robot-Interaction scenario: When a robot is guided around in its future operation area to learn places and objects, the user needs to make the robot follow her to pass through the environment. Additionally, a user and a robot need to position themselves so that pointing and labeling objects and places through spoken dialog is possible. In a study with 22 subjects we observed and recorded movement and positioning during this interaction to understand how robot motion and interaction behaviors can be designed to be perceived as socially appropriate. The questions guiding the study were: how do the spatial distances and orientation of a robot in relation to a human vary throughout a cooperative task in a typical home environment?

We are interested in this spatial management behaviour as it requires the active monitoring and dynamic reaction to each others' movement and position changes. As expectation we want users to feel comfortable and in control of the robot. Additionally, we want to understand how the robot can be enabled to realize the significance of its movements in the context of interaction with a human user. Posture and positioning changes in HRI are therewith prerequisites to read one another's signaling through joint spatial management. It is assumed that it is used in parallel to other communication modalities like spoken utterances. To find the relevant features of such spatial interaction between a robot and a user we let a robot interact with users and analyzed the interaction for variations in distance and spatial orientation.

The remaining paper is organized as follows: The background to helpful concepts from social interaction studies and related research in robotics is given in the next part. In Section 3 the user study conducted is presented in set-up, data-collection, analysis, and findings. Finally, in section 4 a discussion of the findings and conclusions are drawn.

2. BACKGROUND

Findings from such diverse disciplines as archaeology, social interaction studies, Computer Supported Cooperative Work (CSCW), choreography, social and environmental psychology, contribute to our understanding of spatial (inter-) action in co-presence of other people and (interactive) artifacts. Below some concepts and studies found relevant to HRI are given briefly.

2.1 Humans acting in space

The HRI situation of interest here is best described as a *face-to-face* encounter, i.e. a shared physical space enabling synchronous communication exchanges as opposed to remote or technically mediated ones [6]. In such a situation the (nonverbal) communication can be regarded as partly unconscious, meaning that an exchange of signals in such an encounter is not necessarily a choice, but something that can not be avoided – it "simply happens".

Birdwhistell [3] believed that behavior of posture or bodily movements in relation to social and communicational processes can be understood and interpreted as an external visible and observable code which maintains and regulates relationships between humans. Goffman [8] proposed that elements of interactions can be studied to gain an *in situ* natural understanding of events that happen in encounters when people continuously exchange signals of behavior. This would aid the understanding of how "people routinely achieve order in their interactions with one another".

Three kinds of nonverbal communicational, body movement behaviors to be observed in face-to-face encounters were differentiated by Scheflen and Scheflen [19]: The *reciprocal exchanges* of kinesic behaviors are observable body movements or gestures that are displayed by both interaction partners as part of their joint exchange management. HRI still has to find these reciprocal supportive exchanges, i.e. movements by a robot that can be expected to produce certain body movement responses by a user. The *territorial behaviors* allow or prevent the passage of people across a boundary. For HRI this has a direct and important consequence: Studying the territorial behavior in humans interacting with robots could be a way to determine safety requirements on the robot or the robot motion behavior. It is also important to understand for a robot what kind of closeness to human interactors will be preferred in different situations. *Language-orientated* kinesic behaviors that can be regarded as part of the spoken communication are of interest to HRI as they support the verbalization of what is currently being talked about in a discourse between two interactors. A simple example would be a pointing gesture going along with the utterance, "Robot, this is a chair". Without the visual information of the pointing gesture the robot as listener would need to start searching for the object in question.

Humans are trained in social norms, taught to the young members of a social group [2]. For robot interaction behavior it remains an open question which social norms will be established over time. Equally undeveloped are social norms and rules that robots should know and act upon in posture and positioning. Such behaviors in communicative and interactional encounters that are interpreted as orderly are said to be socially appropriate [13], i.e. they are characterized by being perceived as ordered affairs that go mostly unnoticed and are handled without consciously reflecting about them. The opposite are interaction behavior cases where interaction and communication breakdowns occur [9], e.g. by unsuited behavior or spoken dialogue utterances that render a situation as socially impossible.

2.1.1 Hall's Proxemics

Edward T. Hall studied interpersonal distances and coined for his studies the term *Proxemics* [11], i.e. "the interrelated observations and theories of man's use of space as a specialized elaboration of culture" [ibid, p.1]. In the human-robot-interaction context of posture and positioning, mainly three findings are of importance: The classification of interpersonal distances into 4 different classes, the realization of cultural differences in the spatial behavior of people from different countries, and last but not least man's perception of space. From his observations in the US, Hall concludes that social interaction is based upon and governed by four interpersonal distances: (1) *intimate* (0-1.5 feet), (2) *personal* (1.5-4 feet), (3) *social* (4-12 feet), and (4) *public* (>12 feet). The combination of measurable spatial relationships, human ergonomic and kinetic capabilities, different social roles and interaction as well as typical characteristics and interaction situations make Hall's interpersonal distances interesting for HRI. It might be postulated that the most co-present HRI exchanges and reciprocal adaptations between a human and a robot will happen in the *social* and the *personal* distances. The *public* distance is of interest as this seems like an appropriate distance to perhaps try to signal that an exchange can or is about to happen. The *social* and the *personal* distance seem appropriate in theory to facilitate both the communication and the exchange of goods (for example the manipulation with a robotic arm). The *intimate* distance seems to be better suited for exchanges with, e.g. mental commit robots like the seal-robot Paro [21], where touch is an intended interaction modality, resulting in the system giving off heat that can be felt.

2.1.2 *Kendon's F-formation system*

Kendon's F-Formation system [13] is based upon the observations that people often group themselves in a spatial formation, e.g. in clusters, lines, circles, or other patterns. The term *formation* is used to express the dynamic aspect of this spatial arrangement, i.e. the need to actively sustain it during interaction. This often takes the observable spatial form of *small, well synchronized movements* of the participating interactors. An *F-Formation* arises when two or more people form a shared space between them to which they have equal and direct access due to their sustained spatial and orientational configuration. The necessary behavioral organization and movement patterns which are used to sustain this F-Formation is called a *F-Formation system* [ibid, p.209]. The F-Formation system can be applied directly to the interaction encounter between a robot and a human: Between the two a so called *transactional segment* or *o-space*, i.e. a space both to look and speak into, or in which they handle (shared) objects of interest is created and maintained.

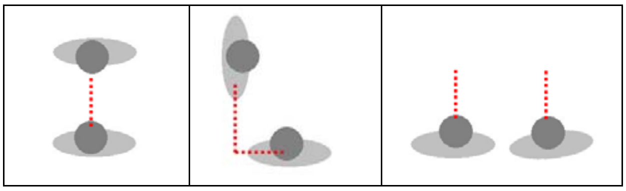


Figure 2: Kendon's F-formation arrangements

Kendon showed in addition that some joint activities and spatial interactions are supported by certain F-Formation system *arrangements*, and thus often are encountered in prototypical situations. In the *Vis-à-vis* arrangement (figure 2, left) two participants normally face one another directly; a *L-Shape arrangement* in which two contributors are positioned so that the frontal surfaces of their bodies fall on the two arms of an 'L' (see fig. 2, middle) usually indicates a joint system in which something is shared in the o-space, e.g. an object of interest. As a last arrangement Kendon mentions the *Side-by-side* configuration (fig.2, right) where the two participants are standing closely together and face the same way. This arrangement is said to occur often in situations were both interactors are facing an outer *edge*, e.g. given externally by the environment in the form of a table, a wall, a kitchen sink, or similar. For HRI it is important to notice that all F-formation arrangements support a *triadic* relationship between the two interactors and an object of shared interest beside a dual relationship between a user and a robot alone.

2.2 Spatiality in interaction with robots

Several systems have been designed or studied to enable the robots to actively use the space in interaction with humans.

Yoda and Shiota [25] take the need for safety in passing a human in a hallway as motivation to develop control strategies for the robot to adhere. Three types of encounters were anticipated as test cases for their control algorithm, i.e. a standing, a walking, and a running person.

Nakauchi and Simmons [15] present another approach by first collecting empirically data on how people stand in line. They use these data then to model a set of behaviors for a robot that needs to get into, wait and advance in a queue for being serviced along with other people there. Butler and Agah [4] varied a robot's movement behaviors and evaluated in a user study how robot speed and robot distance were perceived by users. However, no interactive task between the robot and the user was administered. Equally without a task and need for interaction was the study reported by Althaus *et al.* [1]. They used a complex, room-based sensor array to track the fine movements and spatial adaptation of a group of people and a robot during its initial appearance, its "joining of the group", and finally, the robot's departure again. The study could prove the spatial adaptation from humans and the robot evaluating this adaptation and reacting (in turn) with an dynamic adaptation in positioning.

Prassler *et al.* [17] introduce a robot wheelchair control system that allows the system to stay close to an accompanying person in a busy and crowed subway station, i.e. the robot movement (with a person) in dynamic context is verified. Other people (besides the accompanying person) in this public space are treated as "dynamic obstacles" that need to be avoided as they are limiting the room for navigation.

Topp *et al.* [22] also address the dynamic, joint movement of a robot and its user. However, their robot operation setting is confined to an indoor office space. Additionally the interaction is not only focused on providing a robot navigation component. Instead the authors present a laser-based tracking system that allows the following of users during a so called Human Augmented Mapping mission.

Using Hall's interpersonal distances as defining the interaction, Pacchierotti *et al* [16] devised an algorithm that allows robots to pass people in hallways.

3. USER STUDY

To investigate the spatial interaction of a user showing a robot around we designed a study based upon the COGNIRON idea of a "Home Tour"[5].

3.1 Scenario and Set-up

The trial was based upon a scenario in which a user has got a robot and is ready to use it for the first time. In order to make the robot familiar with the environment it needs to be shown around to learn places and objects that will be of interest to its later operation. To ensure that the robot really has learned these important places and objects, the user is also encouraged to test the robot about the newly learned artifacts and locations.

This was done by encouraging users to send the robot on a "search-" or a "find-" mission to prompt the robot to go back to learned locations or encountered objects. The task embedded in the scenario was thus for invited trial users to (a) get familiar with the robot and navigate it by letting it follow after him or her, (b) teach it places and objects, (c) validate already taught places and objects, and (d) handle interaction practically with a robot, including an initial opening and a closing.

For the study of posture and positioning in HRI the modalities available for interaction and communication play a major role. The robot used in this study is an ActivMedia Performance PeopleBot[1]. It comes equipped with an on-board pan-tilt-zoom camera. Trial users were told that this camera was employed by the robot for object and place recognition. They were also informed that the microphones placed upon the robot were used

---

[1] www.activrobots.com

by the interactive speech system enabling the commanding of the robot by speech.

The trial was conducted in the so called CAS[2] “living room” at the Royal Institute of Technology (KTH) in Stockholm, Sweden. A room five by five meter in size is furnished with IKEA living room furniture, including different tables, a bookshelf, and two sofas (see figure 3). Indicated with numbers are clockwise from the lower left hand corner ❶ the entrance, ❷ the bookshelf, ❸ the Wizard of Oz [10] control station (with a DV-video camera), ❹ a small table with a telephone, ❺ a low coffee table upon which different objects, like a remote control and magazines were placed, ❻ two sofas, ❼ a TV and VCR combination on a small table, and finally, ❽ a small dining table with a fruit bowl, a coffee cup etc.

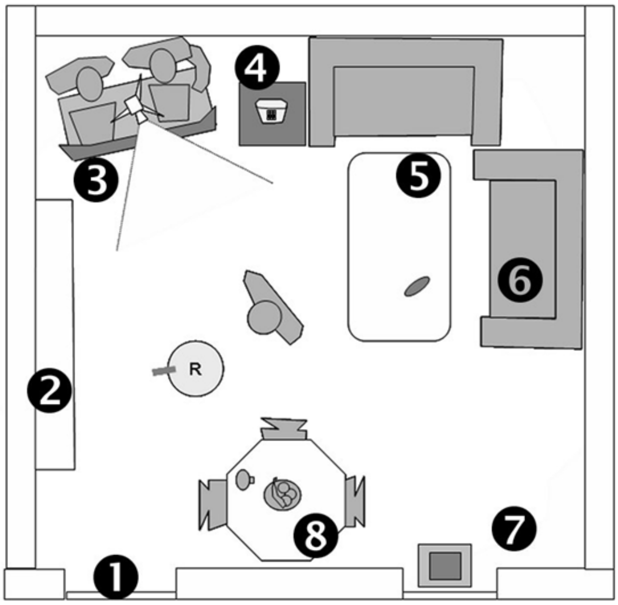


Figure 3: “Living-room” experiment area

The trial subjects were recruited within the Royal Technical Institute of Technology, i.e. young technical students of both sexes. Requirements for selection were that that they did not work or performed research in robotics or computer vision, as this was judged to be the requirement of a robot encounter with novice and inexperienced users. We conducted 22 trials (after 4 initial pilot trials for trial-adjustments) with 9 women and 13 men. Participants of the study were rewarded a cinema ticket for their time and effort.

Upon arrival participants received an introduction to the robot and the task, both in written form and in the form of a short demonstration by one of the experiment leaders. They were then asked to use the robot to teach it new places and objects and validate these. Upon completion of the trial users were asked to fill in a questionnaire before they got debriefed about the simulated nature of the robot’s behaviors in the experiment.

The actual robot behavior was teleoperated by two experiment leaders who used a wireless robot navigation and on-board-camera control for the robot and camera movement and a speech dialogue production tool to interact with speech according to a Wizard-of-Oz research methodology [10].

3.2 Data collection

Multiple data sources were collected during the trial: An external video camera (DV) recorded the trial in audio and video from the experiment leaders’ position and perspective. The room was furthermore surveyed by four webcams running at a frame-rate of about 1 Hz that were placed in the room’s corners in order to ensure that user and robot movements, postures, and gestures would be captured independent on formation within the room The still-images from the webcams are used in the analysis of the spatial relationship and directional setting or formations between a user and the robot (see analysis section below).

Data from a Sick laser range finder on the robot were stored and analyzed with the help of a person tracking system [22]. This data represents information about the spatial distance and positioning of the user under the condition that the user is in a 180° degree half-circle in front of the robot

Two other means of data recording during the trial are mentioned briefly: A system log was saved for all commands that were sent to the robot. Together with the timing information the robot trials can thus be run in a simulator at a later point of time. The different systems mentioned were synchronized against a local Network Time Protocol (NTP) server. Finally, a Marantz digital recorder was used to record the spoken commands on the robot itself for detailed speech dialog analysis and future speech recognition training. This recording is part of the ongoing research cooperation with the University of Bielefeld (Germany) within the COGNIRON project in designing and evaluating of an interactive robot spoken dialogue system [14].

3.3 Data analysis

To find the relevant spatial interaction patterns and ways to categorize them, we carefully examined the data of a few trials first, scrutinized and discussed our methodology before settling on our current approach.

As starting point for the analysis the timeline of the external DV-camera was taken and matched with the transcribed spoken dialogue for synchronization consistency throughout the trial data. Based upon this synchronized transcription the interaction was categorized into three *interaction episodes* termed “Follow” (user guiding the robot around), “Show” (user teaching the robot places and objects), and “Validate” (user testing the taught places and objects by sending the robot on missions to find them again). Another category of interaction was termed “Breakdown”, i.e. scenes where miscommunication and/or task-level incidents led to interruptions in interaction. Often this was accompanied or signified by repair attempts through speech, movement to adapt the position towards the interaction partner, repetitions or a change of interaction strategy by either the user or the robot.

For each of the identified interaction Follow, Show, and Validate-episodes the initial posture and positioning, i.e. the distance, orientation, body posture, gesture(s), utterances, and dynamic changes within the episode itself was annotated. The spatial formation of the user and the robot was analyzed with help of either the laser range finder data or a visual inspection tool (see figure 4). As the laser range finder data is only available when the

---

[2] Centre of Autonomous Systems, www.nada.kth.se/cas

user is standing in a 180° half-degree circle in front of the robot, the visual inspection tool was applied in situations in which the user was standing "behind" the robot or laser data was unavailable.

Figure 4: Visual inspection tool

The visual inspection tool displays different webcam images and supports the annotation of the posture and positioning information. Still images with a frequency of about 1 Hz are first overlaid and fused with a calibration image so that virtual dots on the images mark a grid to calculate distances and positions with. With this aid, marks in the trial environment that could possibly bias users to align themselves with were avoided. Image sequences can be played back and forth and give – together with the time-synchronized transcription of the spoken dialogue the possibility to annotate movements, positioning, and postures.

The outcome of the analysis has been termed a *thick description* giving the literally frame-by-frame commented observations from the trials. These thick descriptions are accompanied with numerically, quantitative interaction episode description (including still image-sequences for illustration) for each of the observed Follow-, Show-, and Validate episodes. This analysis has been conducted for 11 trials so far, i.e. only half of the available data have been taken up for this in-depth spatial interaction analysis.

Focusing on the questions posed initially with respect to the spatial distance and formations of the robot and the human, the following findings section will focus on the results of the spatial management during the Follow, Show, and Validate episodes as analyzed from eleven trial sessions based upon a total of N=321 initiations.

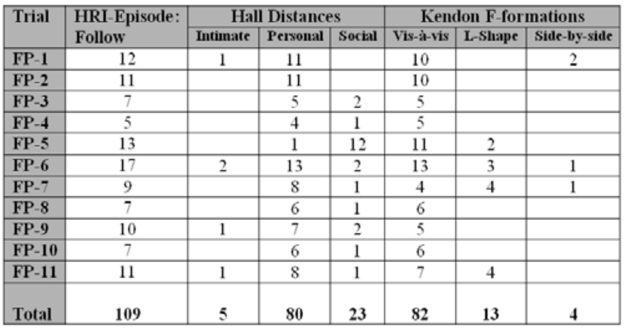
3.4 Findings

Tables 1-3 give the summarized findings for the HRI episodes of Follow, Show, and Validate as introduced above for eleven trial subjects. Column 1 (numbering from left to right) gives the trial-subject's "identity", column 2 holds the absolute number of episodes encountered. Episodes themselves were than categorized according to Hall's social distances of "Intimate", "Personal", and "Social" depicted in columns 3-5. Finally another categorization according to Kendon's F-formation arrangements was tried, giving the occurrences in column 6-8 of either "Vis-à-vis", "L-Shape", or "Side-by-side" F-formations recorded.

Note that the subtotals do not necessarily have to add up to the absolute number of episodes. The reason is that subjects also initiated mission, while e.g. not being in one of the Kendon F-formation looked for. An illustration of such incident is the spatial configuration in the "Validate"-state where users instruct the robot from behind. This is not an F-formation according to Kendon's schemata and thus these cases are missing from the tables.

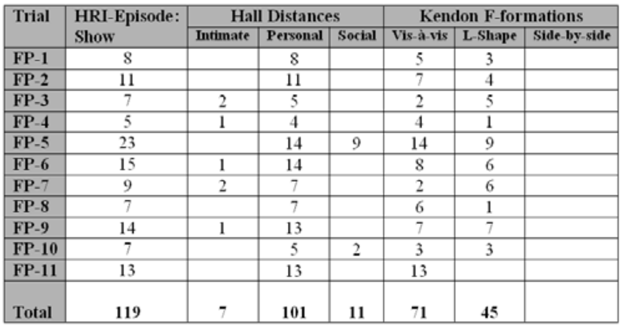
While both the Follow and Show episodes were about equally encountered in overall number of missions, Validate-missions were less frequently observed in general. As the average duration for a Validate-mission is longer than for the average Follow- or Show-mission, only few subjects performed as many Validate missions.

Table 1: Follow-Episodes Analysis



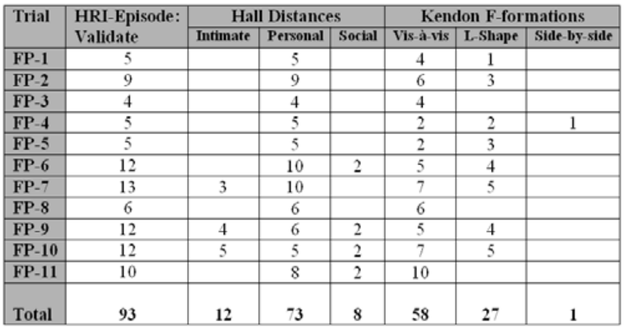
Trial	HRI-Episode: Follow	Hall Distances			Kendon F-formations		
		Intimate	Personal	Social	Vis-à-vis	L-Shape	Side-by-side
FP-1	12	1	11		10		2
FP-2	11		11		10		
FP-3	7		5	2	5		
FP-4	5		4	1	5		
FP-5	13		1	12	11	2	
FP-6	17	2	13	2	13	3	1
FP-7	9		8	1	4	4	1
FP-8	7		6	1	6		
FP-9	10	1	7	2	5		
FP-10	7		6	1	6		
FP-11	11	1	8	1	7	4	
Total	**109**	**5**	**80**	**23**	**82**	**13**	**4**

Table 2: Show-Episodes Analysis



Trial	HRI-Episode: Show	Hall Distances			Kendon F-formations		
		Intimate	Personal	Social	Vis-à-vis	L-Shape	Side-by-side
FP-1	8		8		5	3	
FP-2	11		11		7	4	
FP-3	7	2	5		2	5	
FP-4	5	1	4		4	1	
FP-5	23		14	9	14	9	
FP-6	15	1	14		8	6	
FP-7	9	2	7		2	6	
FP-8	7		7		6	1	
FP-9	14	1	13		7	7	
FP-10	7		5	2	3	3	
FP-11	13		13		13		
Total	**119**	**7**	**101**	**11**	**71**	**45**	

Table 3: Validate-Episodes Analysis



Trial	HRI-Episode: Validate	Hall Distances			Kendon F-formations		
		Intimate	Personal	Social	Vis-à-vis	L-Shape	Side-by-side
FP-1	5		5		4	1	
FP-2	9		9		6	3	
FP-3	4		4		4		
FP-4	5		5		2	2	1
FP-5	5		5		2	3	
FP-6	12		10	2	5	4	
FP-7	13	3	10		7	5	
FP-8	6		6		6		
FP-9	12	4	6	2	5	4	
FP-10	12	5	5	2	7	5	
FP-11	10		8	2	10		
Total	**93**	**12**	**73**	**8**	**58**	**27**	**1**

For the Hall's interpersonal distances it is striking how predominant the "Personal zone" is, i.e. independent upon mission-type did subjects prefer to position themselves in the range of 1 to 4 feet (0.45 to 1.2 meters). Interesting for comparisons is that the number of subjects commanding a Follow-, Show-, or Validate-mission from the intimate zone is much smaller than for example reported in [24] who saw that up to 40% of their subjects approached the robot to a distance of less than 0.45 meters – this figure is much smaller in the results given in Table 1-3 above (e.g. for Follow = 5 (4.6%); Show = 7 (5.8%), and Validate = 12 (12.9%). Although both experiments used an ActivMedia PeopleBot[3] this difference in the number of people coming very close to the robot could not be confirmed in our interaction scenario.

Looking at the Kendon F-formations the dominance of the Vis-à-vis (or face-to-face) positioning of the user towards the robot can be noted, independent upon interaction episode. The "L-Shape" F-formation arrangement is in comparison less often observed. Especially in the Follow-episodes are the L-Shape formations seldom encountered. The Validate and the Show episodes seem more appropriate to be handled in a L-Shape formation – as can be seen from the more frequent occurrences. Especially for the Show episodes, used to present and label objects and places in the environment for the robot, the formation of the L-Shape seems to come natural.

Side-by-side F-formation arrangements were rarely encountered, most often they occurred in the Follow episode. This spatial formation – facing an outer edge together – is likely very dependent upon the environment in which the human-robot interaction is conducted. The set-up in the CAS "living room" might, beside the bookshelf, simply not provide the affordance, e.g. in furniture of this formation to appear very often.

An important limitation to tables 1-3 above should be explicitly mentioned: Each occurrence in the table is based upon a clearly identifiable, often initiated, by speech-dialog interaction episode of Follow, Show, or Validate. It is thus the *starting* point that was taken as marker of the spatial relationship between the robot and a subject. This limits the categorization to a *static* perspective, i.e. the dynamics of change over (even short) time periods is neglected.

The fact that this *dynamic* aspect might however have deeper implications can be seen in figure 5 below. It shows the laser tracking plot of a subject's distance[4] from the robot centric perspective. The user is approaching the robot (coming into the view of the laser range finder) and starts the "Follow-1" episode (depicted through boxes below the graph) after a short while standing still in front of the robot at about 1.2 meters distance. After spoken dialogue initiation the subject takes a step from the robot and waits for the robot's initial movement as feedback (visible as an increase of the distance, then again a decrease). This feedback signal is taken by the subject who starts going towards a corner of the room, rapidly increasing the distance towards the robot (peaking at about 2.3 meters). Arriving at a goal-position the subject stops and turns around waiting for the robot. The robot's approach towards the non-moving subject gives a sharp falling flank at the end of "Follow-1".

Once the robot has reached the subject's position a transition is made by the trial participant into three "Show"-episodes that are initiated after one another without noticeable changes in position from the subject. This is shown through the almost horizontal (distance-) line of "Show-1" and "Show-2".

Note however the small position changes in distance just *before* and at *the end* of the "Show"-episodes (pointed out by arrows in the graph). Almost none-noticeable in the video-data, these small *alignment movements* can be found in the data to often signify transitions from one interaction episode into another. We find these *micro-spatial adaptations* interesting as they might provide in the future a possibility to try sensor-perception-based triggers indicating that new interaction task or episodes are prepared for.

The trial-subject's mission depicted in figure 5 is continued with multiple "Follow"-episodes; the illustration example finally ends with another "Show". While somewhat disturbed by laser-sensor jitter, even the "Show-4" episode is characterized by a straight horizontal line.

From the data we have analyzed so far we saw that different HRI-interaction episodes will also produce different spatial patterns in the sensor readings that monitor the (subject) user's movements and positioning. Summarizing we learned to express the observed dimensions and differences of the interaction episodes of Follow, Show, and Validate by their characteristics: "Follow" is best typified by a *paired-dynamic and user-initiative* driven joint activity which, e.g. can be seen from the dynamics of distance/orientation measurements and high spatial-change frequencies. "Show" has instead a *paired-static, joint interactivity* attribution. Movements are confined to small adaptive and co-operative engagements and each of the interactor can be acting or reacting in shaping the interaction progress. Finally and as in our scenario tried, "Validate" is *neither paired, nor tightly coupled*. Once initiated from the human user the robot is the more dynamic part while the user becomes a supervisor monitoring the progress at best, or possibly, starting a side activity altogether. What becomes even more pressing with this type of interaction episode is thus, how both the robot and the user find back to their *joint* track of interaction at the end of Validate-missions.

4. DISCUSSION

The Wizard-of-Oz controlled interaction of showing a robot around in a singular room and teaching it places and objects to study the spatial interaction behaviour for subjects using a robot for the first time was presented in this paper. Theoretically grounded in work from *human-to-human* social interaction studies, Hall's interpersonal spatial zones and Kendon's F-formation arrangements were tested for applicability in the analysis of human-robot interaction episodes termed Follow, Show, and Validate.

Static measurements showed that Hall's personal distance, i.e. a distance between robot and user in the range of 0.45 to 1.2 meters was preferred by users independent of interaction episode. Furthermore, Kendon's "Vis-à-vis" F-formation arrangement

---

[3] with modest modifications however as far as camera positioning and the presence of a "lifting arm" is concerned for Walters *et al.*

[4] as a graphical reduction, orientation data of the subject was removed;

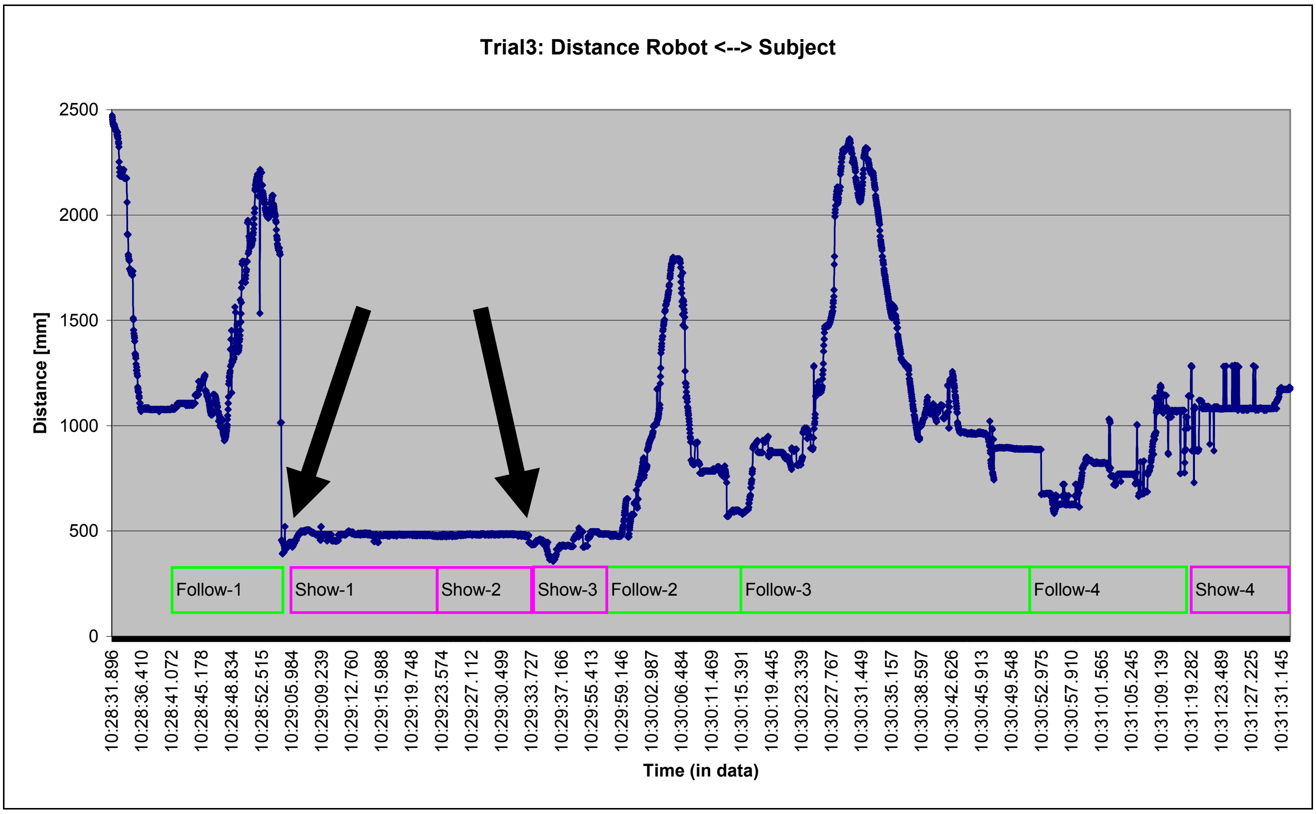


Figure 5: Robot centric laser data plot, showing distance between robot and subject during different episodes

was found to be prevailing among the spatial configurations tested for. A note of caution was raised to the applicability of the terms of both Hall and Kendon however: The dynamic changes and transitions from one interaction episode state into the another one are difficult to express in terms of Hall's interpersonal distances and Kendon's F-formations arrangements when tried in a HRI scenario. It is thus postulated that both Hall's interpersonal distances and Kendon's F-formation arrangements might need to be adapted to suit the dynamics of HRI. A simplistic parameterization barely based on the framework of Hall and Kendon for dynamic spatial relationships and interaction therefore seems unsuited to achieve a socially appropriate robot behavior. It is believed that a more successful alternative resides in the attempt to make other robot interaction components aware both of the communicative as well as the coordinating requirements of spatial interaction. Examples would be to allow the spoken dialogue model to trigger spatial behavior-signaling or pre-emptive spatial movements.

4.1 Future Work

The study reported was constrained in several aspects to keep complexity to a level that allowed us to experiment and investigate aspects of the spatial interaction with a robot. Potential directions to extend our work include gradually phasing out the Wizard-of-Oz control elements with working robot system components. We are also interested in extending the operation area to multiple rooms, and possibly making it necessary to traverse narrow passages together to examine further how elements of the physical environment shape the spatial cooperation between human and robot.

5. ACKNOWLEDGMENTS

The work described in this paper was conducted within the EU Integrated Project COGNIRON ('The Cognitive Robot Companion', www.cogniron.org) and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020. 002020.

6. REFERENCES

[1] Althaus, P., Ishiguro, H., Kanda, T., Miyashita T., and Christensen, H. I., .Navigation for human-robot interaction tasks,. in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, vol. 2, pp. 1894.1900, April 2004.

[2] Argyle, Michael. (1973). Social Interaction. Papperback. London: Tavistock Publications Ltd.

[3] Birdwhistell, R. L. (1952). Introduction to kinesics: an annotation system for analysis of motion and gesture. Louisville, Ky.: University of Louisville Press.

[4] Butler, J. T. and Agah, A., .Psychological effects of behavior patterns of a mobile personal robot,. *Autonomous Robots*, vol. 10, pp. 185.202, March 2001.

[5] COGNIRON. (2003). Annex 1 "Description of Work", EU Project document, Contract number FP6-IST-002020.

[6] Dix, A., Finlay, J, Abowd, G., Beale, R. (1998). Human-Computer Interaction, 2nd Edition. Prentice Hall Europe, Pearson Education Ltd.

[7] Dourish, P. *Where The Action Is: The Foundations of Embodied Interaction*. MIT Press. 2001.

[8] Goffman, Erving. (1967). Interaction Ritual - Essays on Face-To-Face Behavior. New York: Anchor Books

[9] Green, A., Severinson Eklundh, K. Wrede, B., Li, S. Integrating Miscommunication Analysis in Natural Language Interface Design for a Service Robot. *submitted.*

[10] Green, A, Hüttenrauch, H., Severinson Eklundh, K. Applying the Wizard-of-Oz Framework to Cooperative Service Discovery and Configuration. In. *Proc. of the 13th IEEE International Workshop on Robot and Human Interactive Communication,* Kurashiki, Japan.

[11] Hall, E.T. (1966). The Hidden Dimension: Man's Use of Space in Public and Private. The Bodley Head Ltd, London, UK.

[12] Kanda, Takayuki, Ishiguro, Hiroshi, Imai, Michita, and Tetsuo Ono. (2003). Body Movement Analysis of Human-Robot Interaction, International Joint Conference on Artificial Intelligence (IJCAI 2003), pp. 177-182, 2003.

[13] Kendon, A., *Conducting Interaction . Patterns of Behavior in Focused Encounters*. Cambridge University Press, 1990.

[14] Li, S., Kleinehagenbrock, M., Fritsch, J., Wrede, B., Sagerer, G. "BIRON, let me show you something": Evaluating the interaction with a robot companion. In W. Thissen, P. Wieringa, M. Pantic, and M. Ludema, (eds.), Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics, Special Session on Human-Robot Interaction, pages 2827-2834, The Hague, The Netherlands, October 2004. IEEE.

[15] Nakauchi, Y. and Simmons, R.. A social robot that stands in line. In *Proc. of the IEEE/RSJ Intern. Conference on Intelligent Robots and Systems (IROS)*, 2000, 357-364.

[16] Pacchierotti, E., Christensen, H.I., Jensfelt, P. Human-Robot Embodied Interaction in Hallway Settings: A Pilot Study. In *IEEE International Workshop on Robots and Human Interactive Communication (Ro-man)*, 2005, 164-171.

[17] Prassler, E., Bank, D., and Kluge, B., "Motion Coordination between a Human and a Mobile Robot", in Proc. of the 2002 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems, Lausanne, Switzerland, October 2002

[18] Reeves, B. & Nass, C. *The media equation: How people treat computers, television, and the new media like real people and places*. NY: Cambridge University Press, 1996.

[19] Scheflen, Albert E. and Scheflen, Alice. (1972). Body Language and the Social Order – communication as behavioral control. Engelwood Cliffs, N.J.: Prentice-Hall, Inc.

[20] Shepard, R. N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. Psychological Review, 91, 417-447.

[21] Shibata, T.; Wada, K.; Tanie, K. Subjective evaluation of a seal robot in the National Museum of Science and Technology in Stockholm. In: Robot and Human Interactive Communication (RO-MAN 2003), pp. 397-402.

[22] Topp, E.A., Christensen, H.I. Tracking for Following and Passing Persons. In *Proc. of the IEEE/RSJ Intern. Conference on Intelligent Robots and Systems (IROS),* 2005, 70-76.

[23] Turkel, S. *Life on the Screen*. Simon and Schuster. New York, New York. 1995.

[24] Walters, M.L., Dautenhahn, K. Boekhorst, R., Koay, K.L., Kaouri, C. Woods, S, Nehaniv, C., Lee, D., Werry, I. The Influence of Subjects' Personality Traits on Personal Spatial Zones in a Human-Robot Interaction Experiment. In: *2005 IEEE International Workshop on Robots and Human Interactive Communication.* pp. 347-352.

[25] Yoda, M. and Shiota, Y., .The mobile robot which passes a man, in *Proc. of the IEEE International Workshop on Robot and Human Interactive Communication (ROMAN)*, pp. 112.117, September 1997.

Adaptive Systems Research Group, School of Computer Science
University of Hertfordshire

Summary of Human-Robot Spatial Interaction Results

8th September 2005

Human to *Robot* (H-R) approach preferences and tolerance:

- Most adult subjects (approximately 60%) approached the stationary, immobile robot to distances corresponding to Hall's *personal* or *social* spatial zone (0.5m to 1.25m) and the overall mean approach distance was approximately 0.7m.
- Almost all the children took up initial approach distances to a stationary immobile robot corresponding to the *social* spatial zone (1.25m to 3m)
- There is some indication that those subjects with a *proactive* personality (with traits of *impulsiveness*, *creativity*, *general activity* and lack of *shyness*) tend to exhibit greater approach distances to the PeopleBotTM robot.
- A small minority (10% or less) of both adult and child subjects approached to a distance of 2.5m or more, which would be near to the far limit of the *social* spatial zone. In some cases the videos suggest that this was due to a subject being nervous or anxious in the presence of the PeopleBot™.
- A large minority of adult subjects (40%) approached the robot to distances comparable to the *intimate* or *close intimate* spatial zones at 0.45m or less. This may be an indication that the subject views the robot as a lesser or non social entity. It is currently an open issue whether this result would change with habituation (long-term, repeated interaction with the robot), and/or a different (more or less humanoid) appearance of the robot.

Robot to Human (R-H) approach preferences and tolerance:

- Most adult subjects (when standing and not moving) allowed the robot (moving towards the subject in a straight line) to approach to distances corresponding to the *personal* and *social* spatial zones and similar to the results above, with 40% of subjects allowing the robot to approach them right up to the 0.5m robot safety limit.
- The overall mean comfortable approach distance was 0.9m.
- A large majority (more than 80%) also indicated that the robot did not make them feel uncomfortable or threatened when approached closely (to the 0.5m safety limit) by the robot.
- When seated, most human subjects (90%) do not like to be approached by the PeopleBot™ directly from the front. (Note: Of those subjects that preferred a frontal approach, initial observations of the video records indicate that at least some may not have sat straight or faced directly to the front of the chair.)
- Also when seated, individual subjects express a preference for the robot to approach from either their left or right. The actual R-H approach direction preferred overall is slightly biased towards an approach from the subject's right, but in any individual case seems to depend upon an individual subject's preference, perceived practicality, the task and the layout of the area. This suggests that when approaching a seated person, a robot should

assume a default approach direction from the right, but allow this direction to change to the left depending on the individual subject's preference and the physical circumstances.

Comparison of H-R and H-R experiments:

- Generally, comparing the R-H to the H-R experiments, we find a good match of the results. The overall mean comfortable approach distance for R-H was 0.2 m larger than the overall mean comfortable approach distance for H-R (0.9 to 0.7m) which is plausible in the sense that people can be expected to be more cautious when confronted with a moving than a stationary robot. Note: Strictly speaking, the difference is not statistically significant (t-test: Paired Two Sample, P(T<=t) = 0.073) and the conditions between the H-R and R-H approaches differed in that a 0.5 safety distance was used when the robot approached the subject. Thus, we cannot draw any firm conclusions at present.
- The variation among individual subjects is large and supports the approach towards a personalized robot companion, i.e. starting with default settings, but allowing for adaptations.

Note: All results provided above relate to studies performed at the University of Hertfordshire using PeopleBot^{TM} robots. The following publications provide more details. Generalisation across different robot platforms, robot movements, tasks and HRI scenarios is an open issue. However, we hope that the results above provide some constructive input for deciding on default settings related to social spaces in the Home Tour Scenario as part of RA1.