CHRIS
Cooperative Human Robot Interaction Systems

COOPERATION

This document contains:

Deliverable 5 (D5), Month 17 - Systems Engineering Analysis Report

Appendix 1: Snapshot of YARP ports

Appendix 2: D5.1 Adaptive Planning Capability Report (Internal Deliverable)

Appendix 3: Working Paper: Perspective taking in human-robot interactions

All other CHRIS (non-internal) Deliverables can be found on the CHRIS website private document area: http://www.chrisfp7.eu/privatedocs/index2.html

Username: chrisdoc
Password: RN7jq6ZP

Deliverable 5 (D5), Month 17
Systems Engineering Analysis Report

Produced by WP3:
Peter Ford Dominey, Zhenli Lu, Stephane Lallee (INSERM), with contributions from BRL, MPG, LAAS, IIT

Grant Agreement number: 215805

Project acronym: CHRIS
Project title: Cooperative Human Robot Interaction Systems

Project start date: 01 March 2008

Funding Scheme: Small or medium-scale focused research project (STREP)

Date of latest version of Annex I against which the **assessment will be made:** April 2009

Project coordinator name, title and organization:
Prof Chris Melhuish, Director, Bristol Robotic Laboratory (BRL), University of the West of England, Bristol

Tel: ++44(0)1173282539
Fax: ++44(0)1173283960
E-mail: Chris.Melhuish@brl.ac.uk

Project website address: http://www.chrisfp7.eu/

Table of Contents

1. Objectives of D5 and Roadmap

This document is the logical follow-up after D3 Systems Engineering Specification of the System Architecture and Interfaces, month 8, and Internal Deliverable D5.1 Adaptive Planning Capability Report (please see Appendix 2). The framework for this specification is the System Architecture which has been updated based on a merging of architectural specifications in D3 and D5.1.

The objective of this document is to lay out the next level of detail at the system level, in terms of what we will do in the project (i.e. what functions will be implemented in a "superficial/prototype" manner, vs what will be addressed in a deep and robust way). This includes a specification of what specific implementation will be done by the distinct partners.

Background for this document:

In Annex 1, Description of Work (DoW), we specify our method to identify the system functional requirements:

> *"Following meeting 2, Milestone 1, a document is produced, (Deliverable 3, Month 8) detailing functional requirements including the specifics of the overall architecture, and minutia of engineering and cognitive aspects for the project which need to be coherent across platforms. This is following the meeting at IIT, and the Wizard of Oz simulation[1] highlighting the required advances and technical specifications to achieve safe HRI. Initial work towards D3 was performed at NCRM under WP3a, final amendments and delivery of D3 were performed by INSERM under WP3b.*
>
> *WP's 3a and 3b both provide and receive information from WP4, 5, 6, 7, 8.*
>
> *The consortium has already established that, as a priority, an agreed set of 'choreographed' bench-marked interactions must be set. Furthermore, we have agreed that these interactions manifest as the of 'Wizard-of-Oz' simulation in which the robot's interaction with a human is controlled (for example by a human actor or placing the robot limbs or even via limited tele-operation by a human or some combinations). This fixes the constraint envelope for the robot. These enactments specify the interactions and provide a set of comparative benchmarks when the robot has to interact with a human as an autonomous entity."*

D3 was thus successfully delivered. The current document responds to the following specification:

> *"**T3.3** Maintain up-to-date information from all WP leaders on system functionality and specification of both robot platform and systems modules for safe interaction to ensure compatibility (M6-M39) (D5)."*

[1] The "Wizard of Oz" walkthrough was renamed "Scripted Scenario Enactment" (SSE). The decision to perform this scenario enactment was however clearly specified in the DoW from the outset.

Roadmap for the rest of this document:

Section 2 outlines the architecture and the different components. It provides an overview of the functional description of the different components.

Section 3 provides discussion of the scenarios will address, including the definition of scenarios additional to the SSE and in light of evidence from D4 (Experimental data on children engagement in cooperative tasks).

Section 4 specifies the specific implementation contributions of the different partners.

Section 5 places this work in the context of the first annual review and summarizes the original contributions.

2. System Architecture

In the DoW we identified (para 1.3.3, page 21-22) a set of core functions and example scenarios. The principal components of Scenario 1 (tool/object identification), Scenario 2 (holding an object in place), Scenario 3 (Coordination) and Scenario 4 (Task learning 1) (see section 3 of this document for description of these scenarios) were combined into a table building scenario which forms the basis of the Scripted Scenario Enactment (SSE). As planned, the SSE was then deconstructed to identify a set of functions that were mapped onto distinct Work Packages in D3 System Engineering Specification (October 2008). This analysis led in D3 to a specification of the system architecture. Based on the subsequent technical integration work, we have refined the architecture as described here.

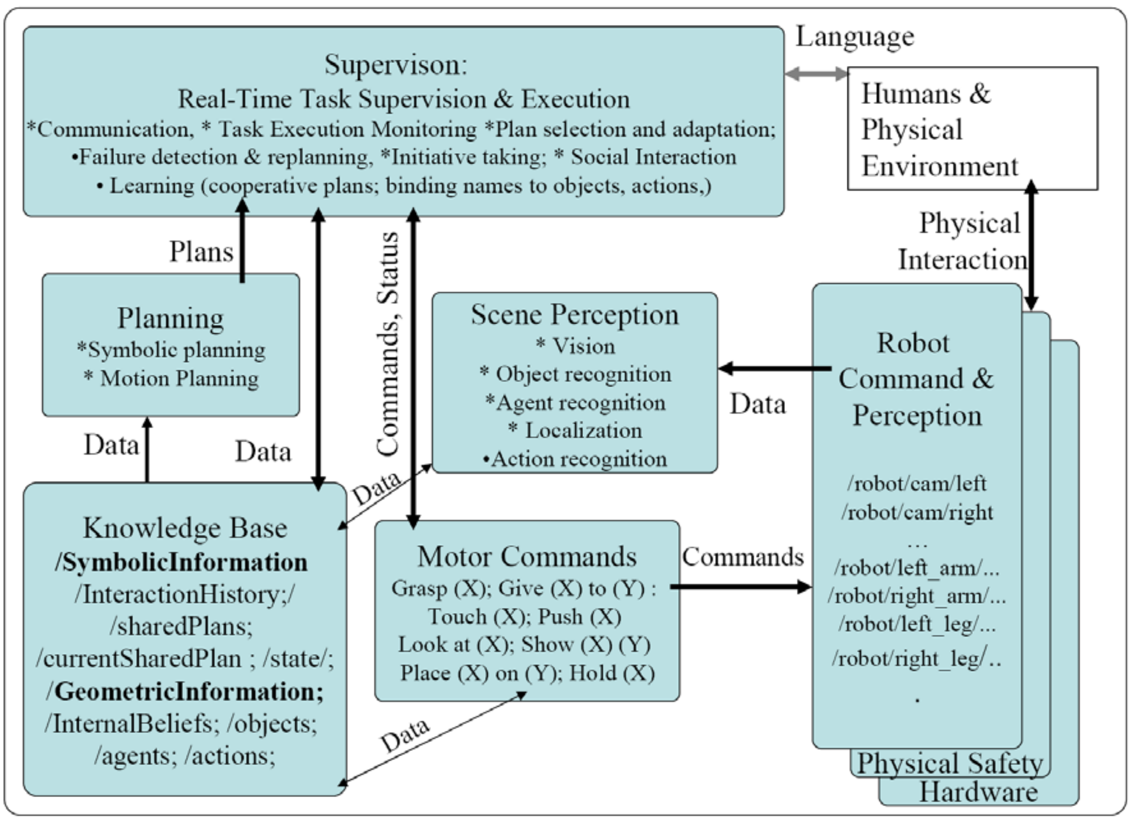


Figure 2.1 : System architecture synthesis, developed based on an analysis of the system architecture specified in Deliverable 3, and in Internal Deliverable 5.1 It corresponds to a system in which the different functional modules communicate via the YARP port interface scheme.

2.1 Robot Command and Perception

This is the robot-specific interface. It provides access to joint actuators; position, force, torque etc. sensors. It provides access to vision and related perception including external motion sensors.

2.2 Knowledge Base

The Knowledge Base embeds and maintains all the knowledge available to the other components. The Knowledge base is to be continuously updated based on

- vision processing related to objects, agents, actions;
- motor command processing
- plan execution
- learning

The knowledge base also defines the set of commands and information requests that can be made to the robot, including high level motor commands and queries (see below sections).

The Knowledge base will form a crucial interface between modules provided by different partners, and thus should be defined as well as possible here. For example, the contents of /InteractionHistory will be used by the Supervision module to allow anticipation and learning, and so we will specify conventions on the contents of what is written to /InteractionHistory.

The Knowledge Base should not be considered as a centralized source of data, but as a distributed information system that includes reasoning. So far we have identified three main components:

1. symbolic information module: it contains symbolic description of the environment, as for example, “the bottle is ON the table”, “the human is NEXT TO the chair”.
2. geometric information module: geometric description of the environment (e.g. object position (x,y,z,theta).
3. Internal beliefs: it contains a description of both the robot’s and the human’s beliefs (of course, the human beliefs correspond to the robot beliefs about the human’s beliefs).

The objective here is to provide a management of the information that will allow the system to synthesize high level actions and to answer to high-level queries in a human robot cooperative object manipulation context. Examples are given here below:

High Level Commands:

1. Grasp (something): Robot reaches towards an object and grasps it.
2. Give (something) to (someone): Robot hands over what it has in its hand to specified person.
3. Touch (something): Robot reaches specified thing and touches it.
4. Push (something): Robot reaches and pushes specified thing.
5. Look at (something): Robot looks at specified thing.
6. Show (someone) (something): Robot shows or points asked thing to a person.
7. Place (something) on (somewhere): Robot places the thing that it possesses onto the specified surface.
8. Hold (something): Robot holds an object still.

High Level Queries:

1. Locations of known objects – “Where is the ball?”
2. Self-attention focus – “What are you looking at?”
3. Estimated human attention focus - “What am I looking at?”
4. Estimated location of human(s) - “Where am I?”
5. Self-behavioral state. - ”What are you doing?”
6. Other behavioral state – “What am I doing?”
7. Known behaviors – “What can you do?”
8. Current behavior – “What are we doing?”
9. Motivation state – “What do you want to do?”

These items have been partially specified in D3 Section 2.3 Memory content specification concerning Objects, Behaviors, and Interaction History. This has also been addressed in D5.1 Section 6 The Knowledge Base (see Appendix 2).

2.3 Scene perception

This level acquires information from the robot/system specific sensors (vision, motion capture, etc.), and performs analysis including object recognition, agent recognition, action recognition.

Interaction with the Knowledge Base: each time a perceptual event occurs (change in object location, action recognition) the Knowledge Base is updated appropriately. For example if a human action is recognized this will include adding a record of this action to the InteractionHistory.

2.4 Motor Commands

This level implements the first level of motor behavior including Grasp (X); Give (X) to (Y); Touch (X); Push (X); Look at (X); Show (X) (Y); Place (X) on (Y); Hold (X). These commands are decomposed into robot specific actions.

These commands will be executed in a dynamic human-robot context and consequently one goal is to define how these commands will be implemented across platforms in order to provide the possibility for the supervision capability developed by LAAS to invoke and control them in a relevant way.

Interaction with the Knowledge base: each time a command is requested and executed, the state of the world, of the task and of the human-robot interaction will have to be updated based on perception but also on data (values, events) produced by the motor commands execution.

2.5 Supervision: Real-Time Task Supervision and Execution

This level is the core human-robot cooperation function, coordinated with Planning and the Knowledge Base. It is defined based on the specification of document D5.1 Adaptive Planning Capability Report. The essential parts of this document are summarized sections 4.4-6 below. (For the full D5.1 document, please see Appendix 2).

2.6 Planning

The symbolic planner corresponds to a task planner. Given the current goal (either individual or shared) the robot has to plan the steps (subtasks) to achieve it. For shared goals, the robot must plan taking into account the human, i.e. which subtasks the human may and should perform. Thus, two streams should be taken into account: robot's and human's actions. Moreover, since we are designing a robot that helps the human, it should try to avoid plans where the human executes most of the subtasks.

The motion planner generates a path from an initial robot configuration to a given final configuration by taking into account robot's kinematics and the environment. Since we are working on a human robot close interaction context, the motion planner should also reason on human's safety. This planning generates motor commands, which are then transformed into motion trajectories by the Motor Command level.

2.7 Safety

A wide range of risks could be considered when humans and robots interact. Even though electrical hazards (e.g. due to worn out insulation), sharp edges and hazardous hydraulic fluids etc. are of real concern when designing a commercial product, physical safety in the context of the CHRIS project relates mainly to unintended collision and the exertion of unwanted forces between the human and the robot. Causes for these events can vary, and include electrical and mechanical hardware failure, software and communication malfunction as well as human mistakes during interaction.

In CHRIS we employ three distinct humanoid robot platforms; HRP2, iCub and BERT2. Amongst these robots, BERT2 is a new platform developed at BRL and consequently allows us to embed sensors (including the data interpretation algorithms) and monitoring software relevant to physical safety (referred to as 'Physical Safety 1' in the illustration below) into the platform. This includes monitoring of the functionality of individual joint drivers and sensors which allows for a gradual decommissioning of malfunctioning parts. Furthermore, the existence of torque sensors on all joints as well as a touch sensitive cover (resembling a rudimentary robotic skin) allows BERT2 to detect impact and interrupt and/or modify trajectories issued from higher level modules via the common motor commands.

Since iCub and particularly HRP2 are at a more mature state of development and access to internal states and sensory data is not readily available, a further physical safety module ('Physical Safety 2') capable of detecting possible hazardous conditions based on external observable signals (reference and real joint trajectory streamed via YARP, human position and human action) is being developed. Robot platforms will provide the appropriate inputs that allow this external physical safety module to stop joint movements at any given time.

Behavioural safety is addressed at the supervision stage (‘Behavioural Safety 1’, BS1) as well as on the level of motor command interpretation (‘Behavioural Safety 2’, BS2). From a supervisory perspective, safety includes the detection of probable task failure due to 'a priori' and experientially learned unsafe interaction and appropriate safe re-planning. Once a motor command has been issued, the joint level robot trajectory will have to be generated, while constantly monitoring the human position, direction of movement, human state of attention as well as providing effective obstacle avoidance. BS2 will intervene and stop the trajectory generator when hazardous conditions occur.

A combined physical and behavioural safety architecture, based on the generic architecture synthesis (figure 2.1), is shown in figure 2.2 below and, in the following text, one example is used to describe how these elements interact.

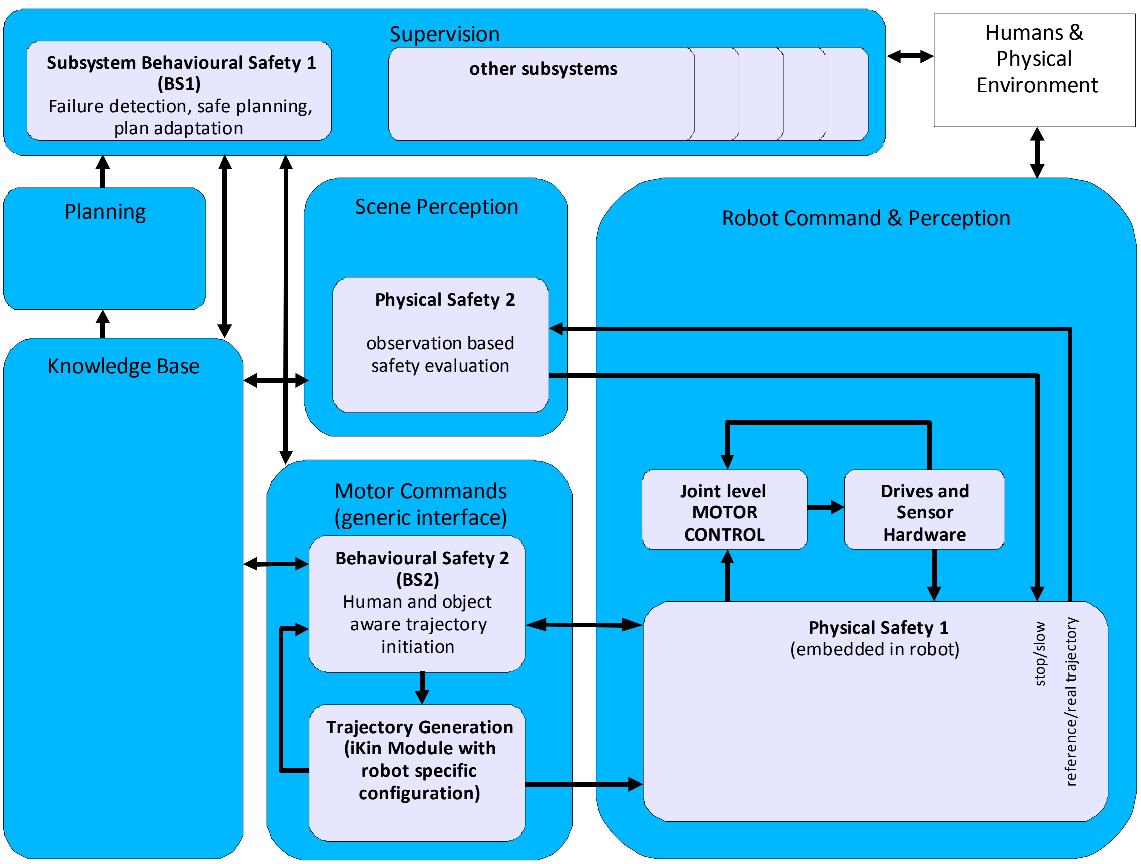


Figure 2.2: Illustration of the system architecture from a safety point of view.

After the supervision system has issued a motor command, and the command has been accepted by the BS1 module, forming part of the supervision module, the generic motor command interface has to deal with this request. BS2 will use the knowledge base to find out about positions of objects and people. Now iKin (an inverse kinematics module supplied by IIT) will be requested to generate a joint level trajectory with a distinct velocity profile. BS2 will use information about the human’s state of attention and the distance between the robot and the human (described using one or more safety indicators) to specify peak velocities and key positions. The trajectory generator streams the reference trajectory to the robot and BS2 constantly monitors the robot's trajectory as well as the human's state in order to interrupt the trajectory generator if the change in interaction conditions (i.e. safety indicators) requires such an action. The embedded physical safety component (Physical Safety 1) of the Robot Command and Perception module monitors the robot's hardware in order to detect possible collisions and/or hardware failure. The physical safety unit is able to stop or alter the joint level trajectory if this is required for safety reasons. Any such action will be propagated back to the behavioural safety modules. In addition, the external physical safety module, Physical Safety 2, (forming part of the scene perception module) is also able to stop the robot's movement based on perceptions of actual joint positions and the trajectory references. Any interference by the physical safety modules will trigger the behavioural safety modules to re-evaluate the motor commands and issue a new trajectory plan. Also, failed movements will be logged in the knowledge base and the supervision system will use this information to identify failure and to adapt future planning decisions.

3. Cooperation Scenarios

Background: From Annex 1 - DoW 1, 1.3.3 Work package overview /List, p 21.

> *Initial Ideas for Scenario Testing:*
> *The scope and elements of the human-robot interactions in CHRIS are set in the initial scenario definitions. This incorporates and sets the balance between sensori-motor control defining the resolution of movements including gesturing and manipulation, degrees of coordination and cooperation with the human. For example, some scenarios can have simplified requirements on sensori-motor skills while maintaining higher requirements on interaction skills and vice versa.*
>
> *In the first phase of the project, the fine detail of the scenarios is defined and the associated behavioural and contingent capabilities of the robot platforms specified. This project benefits from the wide experience of the Max Planck group at Leipzig (WP6) determining the fundamental behavioural and cognitive capacities of human cooperation and role learning by observation. WP6 provides performance metrics derived from human adult-child cooperative interactions which constitutes the basis for scenario specification and later direct comparison. There is an obvious resonance with the expertise in CNRS (WP5) in setting the focus on the adaptive planning and control capabilities that allow 'autonomous' behaviour of the robot engaged in the interaction with a human. The interplay between these themes then sets the agenda for sensori-motor foundations (WP4) and safety for interaction (WP7).*
>
> *Although detailed scenario descriptions cannot be resolved until the project starts, the consortium have already proposed that interactive task definitions be centred around the following generic activities to varying degrees:*
> *Visual identification (WP4, WP5) of tools, limited spoken language comprehension of commands (WP3, WP7), gesture recognitions (WP7), localisation in space for required grasping and trajectory formation (WP5). To explore these capabilities, WP8 will verify technical interoperability. Further than technically working together, however, our modular components create the complementary 'toolbox' to meet features of the envisaged user scenarios, and investigate implications for the future of everyday human-robot interaction; there are many opportunities to play variations on a number of themes such as:*
>
> ***Example Scenarios:***
> *Scenario 1:* *Tool identification and handing to user: The user identifies a tool by naming it, and asks the robot to hand the tool to him/her.*
>
> *Scenario 2:* *Holding an object in place while human works on it: The user asks the robot to pick up an object and hold it in a particular position so that the human can attach it to another part using a nut and bolt.*
>
> *Scenario 3:* *Co-ordinated tool use: The user is obstructed from access either to the tools, or from placing a part in a given location*
>
> *Scenario 4:* *Task learning 1: The simplest tasks will be those that are the most stereotyped, in which the initial and goal conditions are identical from one instance of the*

task to the next. For example, the human asks the robot to pick up a screw from a given box and place it into position in a wooden part to be assembled. This task will combine sensory and motor capabilities of WP4 and WP5 respectively with the sequence learning capabilities of WP5 and the interaction capabilities of WP6 and WP7.

<u>Scenario 5:</u> Task learning2: An extremely complex scenario which may well be outside the timescale of this project would be for a robot to learn involves tasks that may proceed from different initial or final conditions, with the difficulty being the determination of what are the invariant properties that define the goal(s). Furthermore, this could be learned from a third-party observer perspective such that the robot could take the role of either observed participant (robot or human). Certainly, achieving competence in the previous scenarios would provide some of the fundamental building blocks for such a complex interaction.

Each agreed scenario is then de-constructed as an interaction script. For example, for co-ordinated tool use (Scenario 3) where the user is obstructed from access either, to the tools, or from placing a part in a given location one could imagine having to tap a peg into an object with a retaining hole (e.g. piston and gudgeon pin). We specify a set of 'interaction' nodes where, for example, each node in the sequence specifies an element of context, software, speech, sensing capability, interface and work-package requirements.

The table building scenario which constituted the SSE (specified in detail in D3 Systems Engineering Specification Document) takes several of the major components of the scenarios defined above, including Scenario 1 object/tool identification, Scenario 2 holding the table in place, and Scenario 4 task learning. This allowed us to identify a set of core functions, but clearly the project is not limited to these scenario components.

Here we specify additional scenarios which will identify new requirements, and which advance towards complete coverage of the 5 example scenarios defined in the DoW. This is summarized in Table 3.1 below.

Example Scenario 1: Tool/object identification and handing to user	SSE identifying the first table leg and handing to user
Example Scenario 2: Holding an object in place while the user works on it	SSE holding the table while the user attaches the leg(s)
Example Scenario 3: Coordination – helping the user when access is obstructed	Additional Scenario 3.2 Perspective Taking (allows the robot to detect that the user has obstructed access)
Example Scenario 4: Task Learning 1 – rote learning of a task	SSE – learning the leg attachment sequence then applies directly to successive legs
Example Scenario 5: Task Learning 2 – Learning invariant properties from a third party perspective by observation	Additional Scenario 3.1 – Shared Plan from Observation – learning invariant action sequences with ability to to change roles

Table 3.1 Correspondence between Example Scenarios from the DoW, and our current set of scenarios.

3.1 Additional Scenario - Shared Plan from Observation:

"Uncover the target" (INSERM, MPG, LAAS & UWE-BRL)

Contributions from MPG have identified that the system should be able to form a "bird's eye view" of a cooperative interaction, based on observation of two agents performing the action, and should then be able to use this representation to take the role of either agent, demonstrating a capability for "role reversal". This is related to the learning of invariant action properties from the third-person perspective, in Example Scenario 5 from Annex 1 – DoW. Based on these specifications from MPG, we have developed the "Uncover the target" scenario in collaboration with MPG.

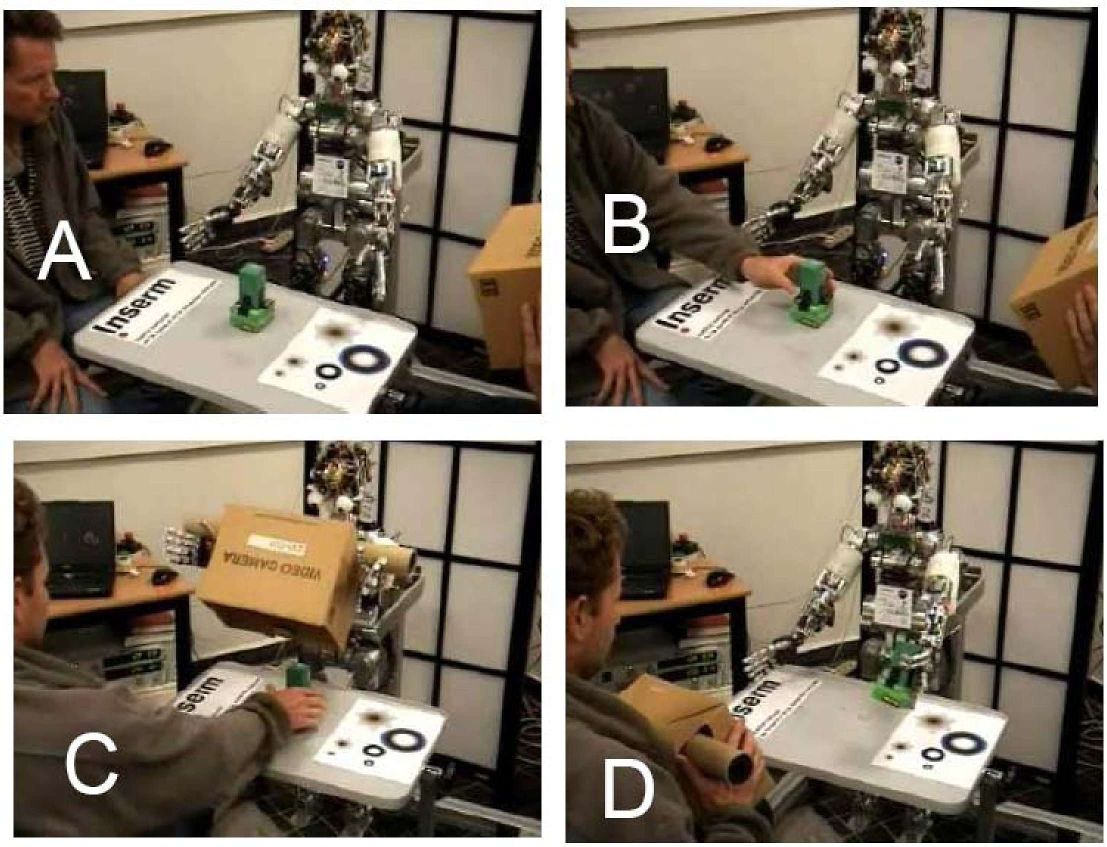


Figure 3.1 Learning a shared plan by Observation.

In this scenario, a two handled box is covering the target object, a small toy. In the demonstration, Larry (the human Left of robot) lifts the box using both hands (frame A, Figure 3.1), allowing Robert (the human Right of robot) to take the toy (frame B, Figure 3.1). The robot should be able to observe the sequence of actions, form a shared plan (i.e. a plan in which actions are attributed to agents), and then use that plan to take either the role of Robert (frame C, Figure 3.1) or Larry (frame D, Figure 3.1) in the cooperative action. This provides a framework for more cognitive learning related to the notion that to be grasped an object must be visible and/or not physically covered/obstructed. The observational learning capability shall extend to any scenario (of arbitrary length) consisting of actions that can be recognized and performed by the robot.

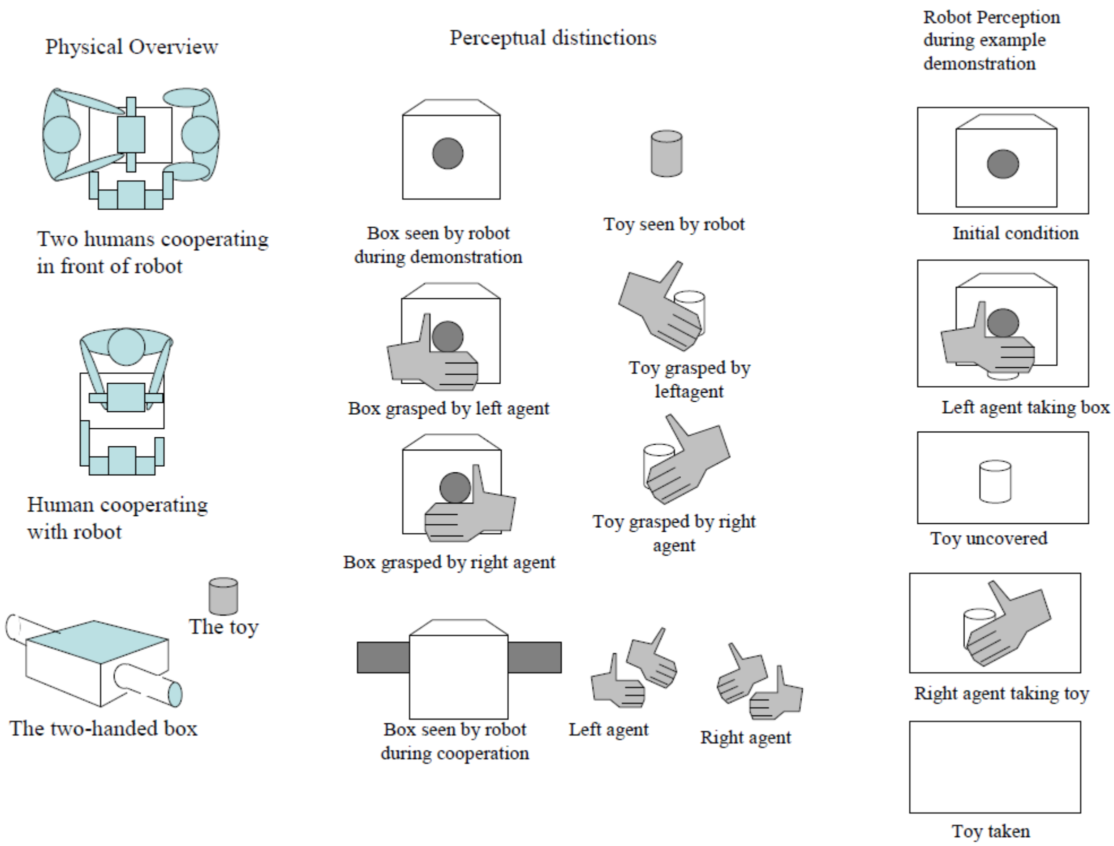


Figure 3.2 Outline of “Uncover the target” scenario. Physical overview. Perceptual distinctions that should be made by the visual system. Robot perception during example demonstration.

Cooperation analysis of the scenario

At a meeting at LAAS Jun 15-17 2009, a cooperative effort between LAAS, MPG and INSERM was undertaken to analyse this scenario within the cooperation framework.

In terms of the CHRIS-Vocabulary, the interaction is characterized by two ***complementary*** actions aimed at a shared goal (retrieving the toy). Moreover, as both agents are supposed to take each role flexibly (lifting versus grasping), they have to be able to engage in ***role-reversal***. In other words, the idea is that they are able to take a “bird’s eye perspective” on the situation, not only representing one’s own action and the outcome of the other’s action in an egocentric way. This is a critical aspect of collaboration (Tomasello et al., 2005). Ontogenetically, it appears to emerge between 12 and 18 months of age, first in simple interactions such as one person holding out a plate and the other person placing something on top – and vice versa (Carpenter et al., 2005). Therefore, the current scenario most closely resembles the role-reversal scenario by Carpenter et al., in which children start to engage at 12 months of age and become more proficient at by 18 months of age. This capacity to engage in role-reversal is also critical for supposedly more complex activities tested in e.g. Warneken et al. (2006, 2007) and the current child studies of the CHRIS-project, namely that by Steinwender et al. (in prep.).

Engaging successfully in the current scenario can be seen as a building-block for the engagement in more complex collaboration as developed in Steinwender et al., as in the box-lifting-scenario the **temporal coordination** of actions is less demanding than in the tests by Steinwender et al. Specifically, in the current scenario it is critical that A lifts the box before B grasps for the toy, whereas in the problem-solving tasks Elevator and Slide (Steinwender et al. in prep), A not only has to perform her action before B can execute hers, she also has to act until B has completed her complementary action by e.g. lifting the elevator and holding it in place until B retrieves the toy or pulling the string of the slide until B retrieves the toy through the hole. Therefore, mutual temporal coordination is necessary.

References:

Carpenter, M., Tomasello, M., & Striano, T. (2005). Role reversal imitation and language in typically developing infants and children with autism. *Infancy*, *8*(3), 253-278.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The ontogeny and phylogeny of cultural cognition. *Behavioral & Brain Sciences, 28* (5), 675 – 691.

Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child Development, 77*(3), 640-663.

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy, 11*(3), 271 - 294.

Steinwender, J., Warneken, F., & Tomasello, M. (in preparation). Collaborative Problem-Solving in Peers

Collaboration analysis for the scenario

Starting configuration: both agents know there is a toy inside the box.

1. Establish **shared/joint goal**
   H: Let's get the toy!
2. R **plans** (gets the recipe for performing the task). Need of **role allocation**. Possible situations:
   a. H goes first.
   b. R goes first **- initiative**. In this case the R would have to "look" for approval.
   c. R asks who goes first.
   d. Both start at the same time: if the robot can detect that the human is also starting, it should get back and then **pick the other role** (since if the robot started already, it’s because it had already allocated the roles).

   To know who goes first, we could use some **social rules**:
   - Hierarchical (chief goes first)
   - First moves, first goes
   - Preferences (I'm better at grasping, so you lift)

   We could do different evaluations based on these.
3. (assuming R goes first) R lifts the box. **Action execution**.
4. H should now take the object. Possible situations:
   a. The H takes the object.
   b. The H doesn't do his job. R should **re-engage**. Possible means:

i. Look back and forth to the H and the goal of the task (the object)
ii. Showing (pointing, moving the object, pushing...)
iii. Ask

5. (assuming H does the job) H takes the object. **Shared attention** on the object.
6. R leaves the box on the table. When does the robot leave the box down? Verify safety aspects.
   a. The toy is not there anymore.
   b. other. See 5. Shared attention to the toy outside the box provides a cue that the object (and thus the human's arm) is not underneath the box anymore.
7. End of the task. Possible means:
   a. Show the object to the other agent.
   b. Place it on a visible position for both.
   c. If not shown after a while, ask.

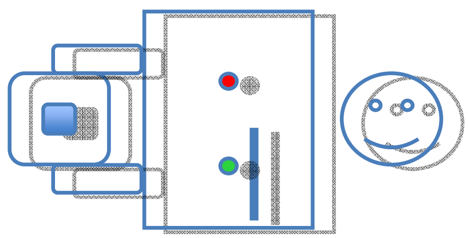
   H's belief should be updated now.

Alternatives:

- To include role-reversal redo the experiment but switching roles. This way we can show that the robot is able of performing both:
- H lifts the box and R realizes that it cannot reach the toy. Then, they change roles.
- Can H and R lift the box together? In this case, we would have **cooperative manipulation**. We need force control.

3.2 Additional Scenario - Perspective Taking (LAAS and MPG)

Based on the working paper by Hamann, Steinwender & Warneken, *Perspective Taking in Human Robot Interactions* (8.06.2009; see Annex 3) in preparation for a CHRIS meeting at LAAS, Toulouse (June 15-17 2009), MPG developed this scenario which addresses aspects of the perspective taking required for Example Scenario 3 from Annex 1 - DoW. From the discussion with MPG we will implement perspective-taking abilities in the robot (especially perspective taking level 1) and we will test them in the following concrete scenario:



In the below extract R refers to the robot on the left and H refers to the human on the right:

"Two balls are on the table, both visible to R. H asks the robot for the ball. Without level 1 perspective taking skills, the situation would remain ambiguous, and R again would have to ask. If, however, these skills exist, R can infer that H refers to the red ball from knowing that the barrier prevents H from seeing the green ball. Hence, it should hand the red one over (cf. Trafton et al., 2005). Conversely, if H does not ask R to give him the ball but instead looks for a ball and asks for help, R should be able to infer that the ball H can see is probably not the one he is looking for, and thus should give him the green one (cf. Moll & Tomasello, 2006). Note that the two different questions evoke two different actions, but the same level 1 perspective taking skill is necessary to solve either." (Hamann et al., 2009,)

This study will be conducted in a way to build a set of perspective taking abilities for a robot in a face-to-face situation. One instance can be the LAAS HRP2 "table and chairs" set up. The ability will be based on a core geometric reasoning library (Marin-08).

References:

Moll, H. & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology, 24*, 603 - 613(11).

Trafton, J., Schultz, A., Bugajska, M., & Mintz, F. (2005). Perspective-taking with robots: experiments and models. In *IEEE International Workshop on Robot and Human Interactive Communication, 2005. ROMAN 2005,* (pp. 580-584).

Hamann, Steinwender & Warneken, *Perspective Taking in Human Robot Interactions* (see CHRIS Wiki: Supplementary Documents Page: http://fp7chris.pbworks.com/Suplimentary+Documents)

L.F.Marin Urias, E.A.Sisbot, R.Alami Geometric tools for perspective taking for human-robot interaction, Mexican International Conference on Artificial Intelligence (MICAI 2008), Mexico, October 2008, 11p.

3.3 Additional Scenario - Face to face scenario (LAAS, MPG)

LAAS is building a "game" scenario that will mainly show the following capabilities in the context of joint cooperative activities:

- Establish joint goal
- Commitment to the joint goal
- Mutual responsiveness
- Perspective taking for shared attention
- Mind reading and activity recognition
- Re-engagement
- Role reversal
- Multimodal communication

This experiment will benefit and build upon results obtained in experiment 3.2 and the study that will performed in the framework of 3.1. Specific task and motion planning capabilities will be "added" here in order to provide a framework for more elaborate management of knowledge as well as for performance evaluation and "longer" human-robot interaction.

We will perform experiments to evaluate the robot's performance based on the following criteria:

- Task performance: alone or in cooperation (individual vs. shared goal)
- Knowledge of the task: completely or partially known
- Roles: expert/novice or team-mates (peers)
- Task efficiency/acceptability: if several plans exist, selecting the most appropriate based on different characteristics (socially acceptable, efficient,)

The scenario will include a human in front of HRP2 and probably Jido moving around. While the main interaction will be focused on the human and HRP2 (who will be seated), Jido could perform navigation tasks in order to get or take objects from other places in the room.

We will use motion capture for human detection and tagged objects for object recognition. We are currently working on the details of the scenario.

4. What will be done, and who does what

In this section, we present a coherent overview of the work to be done, where the CHRIS partners specify what they are doing in the context of each of the functional components of the system as identified in Section 2. Please note, section 5.2 will summarise the commercial off the shelf (COTS) vs. original contributions.

4.1 Robot Command and Perception

INSERM: Will use the YARP interface to iCub ports for position and velocity control of joints, and for access to joint position and velocity, and access to the stereo camera system.

IIT: will implement force control for the robot iCub, the interface to the robot will be extended to give the possibility to 1) control the force exerted by a given joint 2) tune the stiffness of the robot.

BRL: Will use a YARP trajectory streaming interface for position and velocity control of the 28 joints. BERT2 will stream its position and torques at 50Hz for each joint. Pressure on selected parts of its shell will also be available as a YARP stream. A health update of the robot is also available as a further stream. BERT2 is also equipped with an expressive face based on a 3D animation and has several YARP streaming interfaces to animate the eyes, eyelids, eyebrows and the mouth. Furthermore, BERT2 provides a high level interface to produce expression such as happy, tired, surprised etc. and a further input to open the mouth which can be employed to animate speech production.

LAAS: Will extend their existing functional modules based on GENOM and build new ones to offer a set of lower perception and control primitives to our robots HRP2 and Jido for scenarios 3.2 and 3.3. While technical challenges are still to be tackled, this is not to be considered a scientific contribution to the project but a technical contribution.

4.2 Scene perception

IIT: Will provide a generic YARP-like port interface to vision processing. Vision processing will be provided by the Spikenet vision system. The current functions are specified here.

virtual int	getObject (const std::string &object, ImagePositions &pos)=0
	Get the visual (or cartesian) position of an object.
virtual int	getAll (Objects &o)=0
	Get the list of the objects identified in the image.
virtual int	getObject3D (const std::string &object, ImagePositions &pos)=0

Get the 3D position of an object, compatible with iKin.

INSERM: Will use these interfaces for accessing Spikenet vision software for template-based object recognition. They will cooperate with IIT in the implementation of these interfaces.
Object recognition: In all scenarios, templates for required objects will be prelearned. This provides a reliable and robust vision capability.
Agent recognition: will be based on hand recognition. During cooperative action demonstration, agents will be to the left and right of the robot during action demonstration, and the hand configuration will be used to distinguish between two agents.
Action recognition: will be based on object displacement, and co-localization (contact) between agents and objects (specified in Dominey & Boucher 2005). Actions will be specified as an <agent, action> pair, where actions are of the form specified in section 2.6.1.
IIT: will work on mirroring behavior and action understanding. The robot will use previous experience to identify actions performed by an external agent (similarly to: Fitzpatrick et al. 2003).

References:

P. Fitzpatrick, G. Metta, L. Natale, S. Rao, G. Sandini. *Learning About Objects Through Action: Initial Steps Towards Artificial Cognition*, In 2003 IEEE International Conference on Robotics and Automation (ICRA). Taipei, Taiwan. May 12-17, 2003.

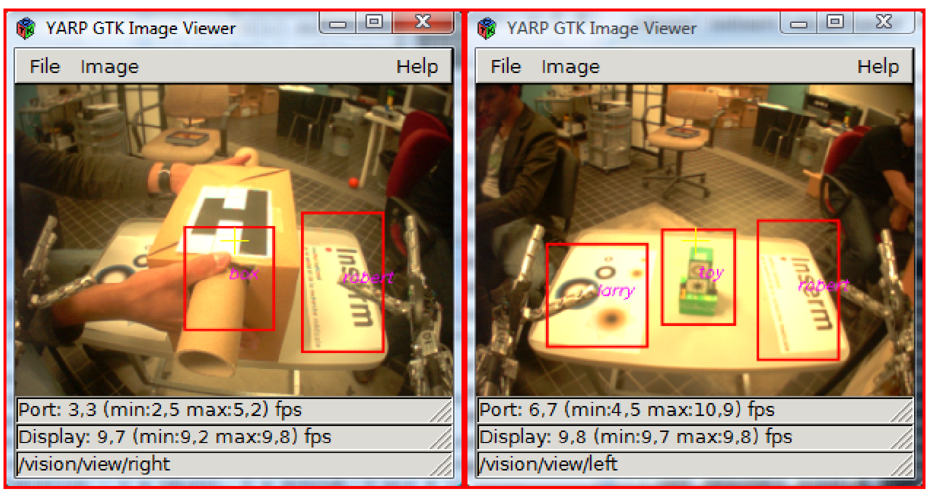


Figure 4.1 Example vision processing seen from iCub's right eye. Left frame, the users hand covers the geometric pattern on the table, and thus blocking vision of the pattern. This is used to attribute agency to the user on the left. Right frame, recognition of the two agent patterns, and the toy object on the table.

BRL: Will use the VICON motion capture system (comprising 8 stationary IR cameras and including IR illumination) using light reflective markers arranged into unique patterns, to distinguish between scene objects and their orientation. This provides reliable and robust 360 degree vision. The human interacting with the robot will be wearing a garment that is also equipped with such markers, thus making body positions and postures available to the system.

BRL will develop both low level and high level software for scene perception using motion capture. The lower level software, the 'VICON-YARP link', will translate motion capture data of a human and objects (e.g. positions, rotations, visibility flags) into a set of YARP ports. As the modules accessing those ports are not fully specified, the ‘VICON-YARP link’ will be designed with maximum flexibility in mind, i.e. port names, parameter names and content order will be changeable.

The ‘VICON-YARP link’ ports are going to be read by at least two software modules, one of which will communicate parameters into the knowledge base through a common interface, and another that will deduce gestures (e.g. pointing) from the stream of changing positions and postures.

These modules constitute our implementation of 3D interaction scene perception using motion capture, and will be available to all partners. For example, our VICON interface will be able to make use of the VICON system at IIT and consequently can be used to enhance iCub’s sensing capabilities.

The higher level gesture detection module will be developed in the context of both ‘VICON style’ 3D motion capture data and also camera based systems, assuming that the required anatomical features can be tracked from the video stream. In this way other partners, whose robotic system does not have motion capture, will be able to use the gesture detection module.

BERT2 will also be equipped with a high precision gaze tracking system and we provide a stream of the head and eye rotation of the human interacting with the robot if he/she is in the robot’s narrow field of view.

LAAS: Will use simple tools for perception in order to provide real-time sensory data to the system. They should not be considered as scientific contributions to the project.

- Object and Agent recognition: a module called VIMAN will be implemented which will allow detecting, recognizing and tracking tags in 3D.
- Adaptations of a motion capture system that will allow providing head and arm motion of the human in front of the robot.

The main contribution here will be the development of a sensor-based geometric reasoning system that will allow the robot to have a perspective-taking ability. This will be refined and incrementally developed in the framework of scenario 3.1 then 3.3.

4.3 Motor Commands

INSERM: Will use the high level commands defined by IIT below. In this context, actions will be executed taking into account object-specific constraints. That is, given different sizes of objects, the Grasp, Hold and Place-on actions will be executed for specific objects, coordinated by the Knowledge Base. The Knowledge Base will include grasp (action) parameters for different known objects. INSERM will initially use a combination of vision guided, and pre-calibrated position information, and will gradually transition to more complete vision guided action.

All command requests pass via the Supervision to determine if they are defined and executable. INSERM will coordinate in order to establish how the supervision capability developed at LAAS can be used on different platforms.

Interaction with the Knowledge base: each time a command is requested and executed, the Knowledge base is updated, including adding a record of this command to the InteractionHistory.

IIT will implement modules to convert High Level Commands (Grasp, Give, Touch…) into motor commands that can be understood by the platform specific Robot Command & Perception module (joint level velocities or positions generated by iKin). These modules might receive input from the Supervison which can provide suggestions/requirements on how to plan the motor action (e.g. avoid specific obstacles and/or use a certain stiffness). These modules provide YARP port interfaces to access the robot-specific implementations of these different commands. Current commands include:

virtual int	grasp (int hand, const std::string &object)=0 *Grasp an object with a given hand.*
virtual int	release (int hand)=0 *Release an object in a given hand.*
virtual int	point (int hand, const std::string &object)=0 *Point a given hand to an object.*
virtual int	reach (int hand, const std::string &object)=0 *Reach for an object.*
virtual int	orient (int limb, const std::string &object)=0 *Orient a limb.*
virtual int	goTo (const std::string &location)=0 *Go to a specified location.*

The motor interface that we have defined requires only limited information including the name of the object to be reached/grasped or towards which the robot should orient itself. This is not enough to plan the action. The missing information includes: the list of possible grasp points and corresponding hand orientations. This will be retrieved by the motor command module from the knowledge base (using the memory interface). Obstacles and/or environment model are not required at this level. We understand that the generation of more complex behavior might require more information (for example the purpose of the grasp or the particular affordance to exploit). Such constraints will be defined later on in a more advanced interface.

IIT will not plan to represent the environment geometrically. Vision will provide what is needed, we plan to cope with unexpected obstacles using force control.

These motor level commands will avoid dealing with obstacles explicitly, at least initially. The robot has a limited workspace and fine planning to avoid obstacles is probably hampered by our ability to detect them by vision (this means simply that if the object is large enough, then the robot can't reach behind it, otherwise it's perhaps difficult to detect). We can use vision to decide whether to start a reaching/grasping action and force feedback to detect contact with the environment. The grasp/reach functions could return an error code in case the action fails because of an unexpected obstacle was encountered (error codes to be extended).

The supervisor will interact with the motor command module using the interface we have defined. Work in progress will identify how geometric reasoning in supervision and planning can augment the motor level command capabilities.

BRL will implement our own platform specific interpretation of the high level motor commands (Grasp, Give, etc.) consistent with these system level definitions, and will implement some behavioural safety strategies at this level. We will employ iKin from IIT and take into consideration the human's position and state of attention to generate joint level trajectories. The behavioural safety unit (BS2) will be in constant contact with the embedded physical safety at BERT2 in order to ensure safe re-planning if collision avoidance or recovery is necessary.

LAAS: will have two activities in this framework. The first involves the implementation of a set of motor commands to run the planned experiment on HRP2 and Jido. The second activity is the design, together with IIT and BRL on the modalities that will allow the robot supervision capability (to be developed by LAAS) to invoke and control the generic motor commands developed by IIT and BRL in a relevant way.

4.4 Supervision: Real-Time Task Supervision and Execution

INSERM: In the context of the Supervision, INSERM is concerned with the acquisition and execution of new shared cooperative plans. INSERM will implement the following capabilities:

Execute action: when a command such as Grasp(robot, bottle) is requested, the Supervision will verify that the grasp command is defined, and that there are grasp parameters defined for bottle, and will take corrective action if these elements are undefined.

Learning by Spoken Language Programming: Here the user makes spoken commands, indicating what the robot should do, and what he/she (the user) will do. A preliminary version of this has been implemented in Dominey et al. Humanoids 2008. To the extent that the human actions can also be executed by the robot, role reversal and helping can be done.

Learning by Observation: Here, two users demonstrate a cooperative activity as specified in D4 from MPG. Specifically, the demonstration of the current scenario corresponds to ***demonstration stage 2*** of Steinwender et al., i.e. a social demonstration in which two agents perform both actions successfully until the goal is achieved. This is the level of detail that children at 1.5 to 2.5 years need to engage successfully in novel collaboration tasks. The system segments the visual scene into discrete actions attributed to the two agents. When necessary the robot requests clarification. This relies heavily on Scene perception identified in 4.2.

Engage in Learned Cooperation: Once a shared plan has been committed to the Knowledge Base, it can then be executed. INSERM here provides a skeleton capability that simply monitors the execution of the human and robot behaviors in the correct sequence. (INSERM anticipates that **CNRS** will provide much richer capability here.)

Edit learned plans: if the user determines that there are errors in the learned plans, a simple spoken language editing capability can be used to modify, add or delete a given action in a plan.

Learn object name: If an action is requested that involves an object name that has not been assigned by the system, then the user will be offered the opportunity to show the system the referent of that name, by pointing.

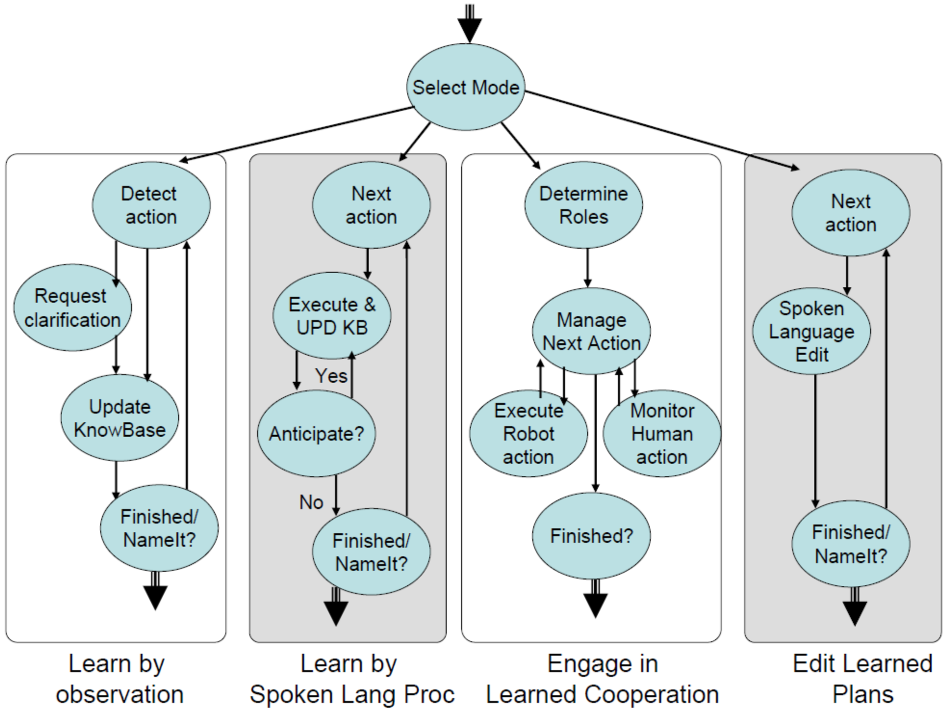


Figure 4.4 Functional Architecture for acquisition and execution of new shared cooperative plans

LAAS: we will design, in the next period, the internals of the supervision system (task execution, joint task management, initiative taking…). From this analysis we will be able to detect which components can be and should be integrated within the other robot platforms.

One key aspect is the identification and categorization of the abilities involved in shared activity, based on interaction with the other CHRIS partners. This very ambitious endeavor is limited here to a context of close human and robot manipulation activities.

In general, supervision is composed of two main elements:

- Contextual task selection and refinement: The supervision system plans and refines the task based on the current constraints of the context.

- Adaptive task monitoring: Based on the progress of the task performance, the supervision system should adapt the monitoring strategy.

In addition to these two, we should include the following components since we are dealing with human robot cooperative interaction:

- Initiative taking and intentionality: We want to provide our robot with an explicit treatment of motivation that will drive it to act, to disambiguate perception to understand human intentionality, and to engage in cooperative tasks. This motivation ability will be combined with a planning ability allowing to determine its capacity to act and to help in a given HRI context.

- Cooperative HR task achievement: According to the type of task -individual or cooperative- and to the human behavior, the supervision system should select and refine the required subtasks in order to achieve a shared activity.

- Adaptive communication for human robot interaction: Based on common ground, beliefs about human attention and preferences, the robot should choose the most appropriate way to convey an information when needed.

4.5 Planning

LAAS will use a Human Aware Planner for planning the task to execute. Its main features are:

- use of a **temporal planning framework** with the explicit management of two time-lines (for the human and the robot),
- a **hierarchical task structure** allowing incremental context-based refinement
- a plan elaboration and selection algorithm that searches for plans with **minimum cost** and that satisfy a set of so-called **"social rules**".

The main characteristics of this high level planner have been described in D5.1 (see Appendix 2). In addition, in the framework for scenario 3.3, LAAS will build a set of geometric reasoning capabilities to allow motion planning for the robot.

Task execution in Co-operative Human Robot Manipulation: LAAS are in the process of developing a set of elementary actions (hold, touch, point etc.) and basic tasks (grasp, pick_place etc.). Generally the basic tasks will be executed as a sequence of elementary actions. The design is done keeping in mind the following points:

- Implementation-Oriented definitions.
- Some or all arguments may not be provided.
- It should be robust to extract the missing arguments.

The LAAS approach is:

- A set of complementary dedicated functions will be provided for every argument, which will decide the best value for missing argument for a task or elementary action.
- Each such function will evolve and enhance during the implementation of the entire system.

For example if the human asks to the robot "Put the glass near me", the proposed approach will execute this command as follows:

Through access to the knowledge base, a candidate ***put()*** commands can be selected. At this level no other direct information will be known to robot for executing the task put(object, from, to). Then a set of complementary functions associated with each argument will be executed. After that the set of elementary actions will be called to execute the task put(). In those elementary actions again if some arguments are missing, the corresponding set of complementary functions will be called.

The choice of the parameters and the execution modalities will be performed though access to motion and task planning processes that explicitly take into account the presence and the current activity of the robot human partner. Figure 4.5 describes a preliminary design of the entire control flow.

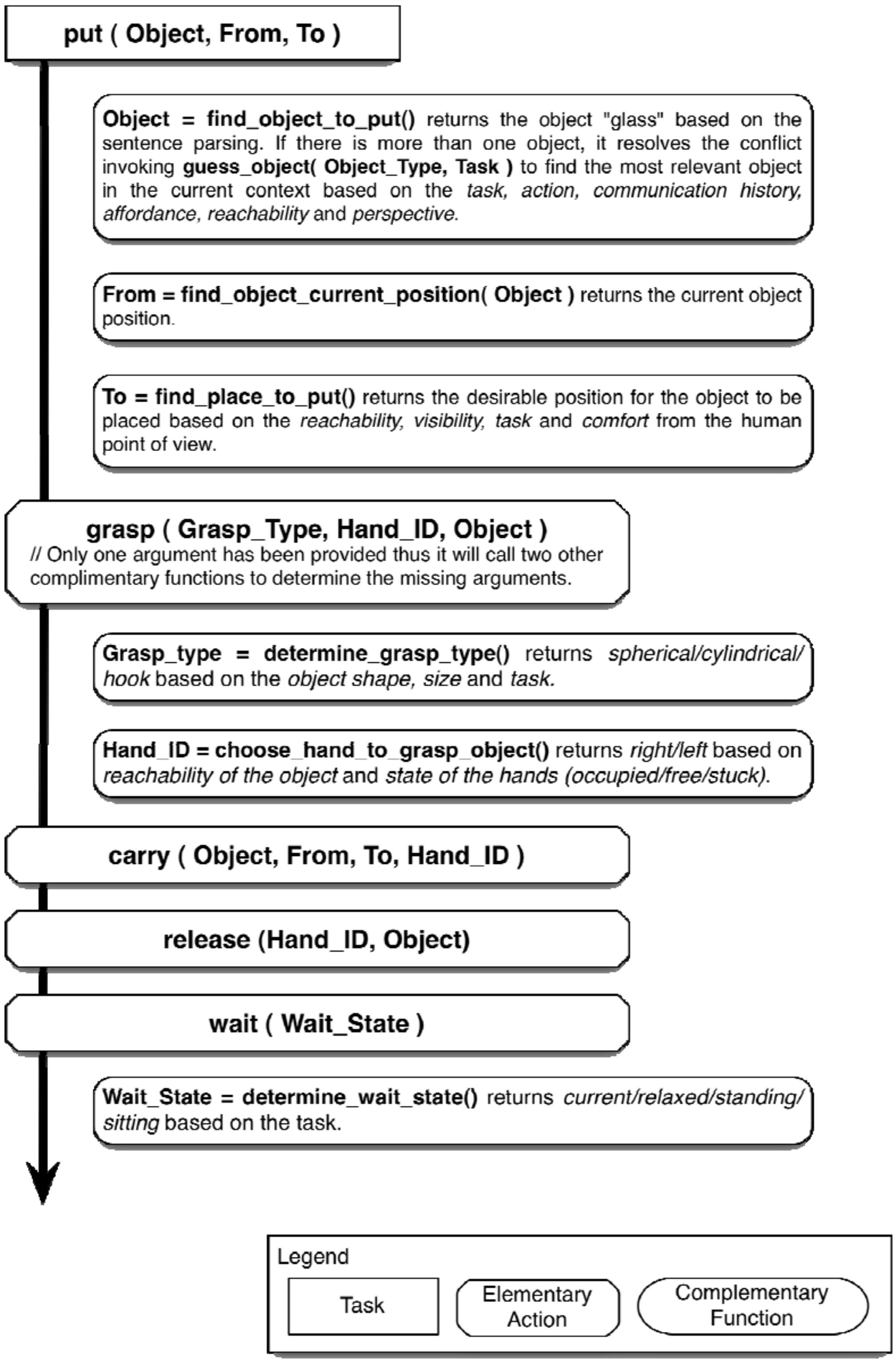


Figure 4.5: Action refinement in a HR context.

4.6 Knowledge Base

INSERM: will collaborate with IIT and LAAS in the implementation of memory capabilities as specified in section 2.3 including state of known objects and behaviors, and the interaction history which is a literal sequential record of actions performed by the human and robot during their cooperative interaction. INSERM will also implement a shared plan representation as an ordered list of actions in the form <Agent, Action, arguments>

IIT: will implement the YARP-like port interface to the memory component of the knowledge base. The current set of functions are based on the specifications from WP3 and include the following:

virtual int	getObject (const std::string &name, ObjectDetails &d)=0 *Check if an object is known.*
virtual int	getAllObjects (StringList &names)=0 *Return all objcets known to the system.*
virtual int	getBehavior (const std::string &behavior, StringList &actions)=0 *Get a behavior known to the system.*
virtual int	getAllBehaviors (StringList &behaviors)=0 *Return all behaviors known to the sytem.*
virtual int	setObject (const std::string &name, const ObjectDetails &d)=0 *Add an object to the memory.*
virtual int	setBehavior (const std::string &behavior, const StringList &actions)=0 *Add a behavior to the memory.*

LAAS will work on the description of the contents and the engine of the Knowledge Base. Feedback from the partners is essential to take into account the needs of the overall system. Particularly we would like to identify the key "data-structures" to be shared among the rest of the robot components (e.g. shared plans, interaction history…).
LAAS will provide the rest of the partners (if agreed) the symbolic information module and the internal beliefs module (described in Section 2.6.). Both sources of information will be implemented using an ontology.

The LAAS OpenRobotsOntology Server: The knowledge base will be provided by LAAS as a server for cognitive application (the OpenRobotsOntology Server).

The server will actually wrap an ontology (the OpenRobots Ontology or "ORO") that acts as a knowledge base and will add both remote access methods and cognition-related capacities in relation with the ontology (cf. figure 4.6).

Knowledge Base Overview

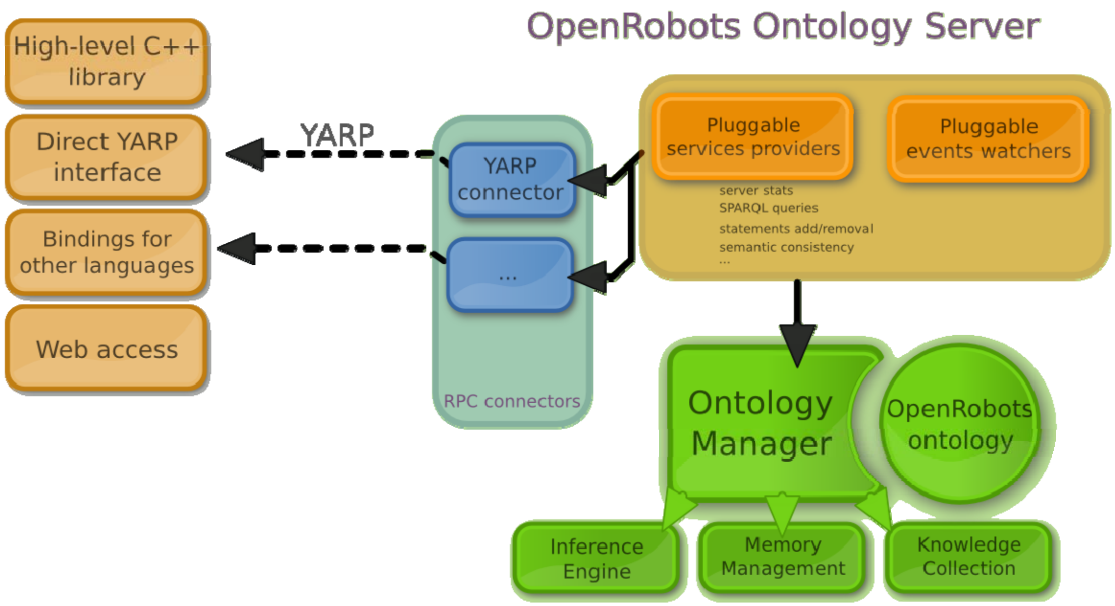


Figure 4.6 The Ontolgy server and its interfaces

The ontology, as a formal specification of a shared domain conceptualization [Gruber1993], aims to be the common store for all the robot's knowledge (both a priori and acquired) and beliefs. It can provide a shared system of symbolic representation whose semantic consistency may be checked.

In the context of CHRIS, the main conceptual areas to be covered by the ontology are:

- Robot's own beliefs and state, as well as models of other agents beliefs and state.
- Manipulated objects, including geometric data.
- Plans, intentions, interaction history.

Chunks of knowledge are represented in the ontology by statements (a triple [subject-predicate-object]) that are contextualized (i.e. connected to each others) and with possible associated meta-statements (this enables meta-cognitive functions like memory, see below).
The ontology, as a formal base of such statements, can be processed and offers support for inference (first-order logic and rule-based inference) from within the server.

The server will be open-source and written in Java. It will rely on an external library (Jena or OWLAPI) to load and manage the OWL ontology. On the top of this, the server is expected to offer convenient methods to remotely access the ontology backend. Supported transport mechanisms will include at least YARP but the server is expected to be easily extended to other protocols.
To ease the development of client modules, a C++ library will be written to allow both low-level and high-level interaction with the ontology server. Monitoring tools (like on-line visualization of the ontology) are planned as well.

The server design is geared towards extensibility: external RPC mechanisms (like YARP) and the ontology backend will be bound by so-called "services" that can be easily added. The first versions of the server should ship with standard services to directly access the ontology (ontology queries, addition and removal of statements, check of semantic coherence, check of facts...) plus some server monitoring functions.

The server will also offer cognitive triggers mechanisms: when a specific event is asserted or can be inferred (e.g. a new object is discovered or a human is close to an object or a class of objects), the server should be able to trigger external processes through pluggable "events watchers".

Other specific developments planned within the frame of the CHRIS project include a framework for "knowledge collection" and a generic model of memory.

By "knowledge collection", we want to address the difficult issue of concept fragmentation: for instance, a vision module may perceive and instantiate in the ontology a new object. Later on, a human may refer to this same object in a sentence. The robot's cognitive framework must have means to match the two concepts. This issue sums up often to a grounding issue, but not always.

Concerning memory model, we would like to conduct research on the lifespan of asserted statements based on the class of memory the statement belongs to (working memory, episodic memory, etc.), and how reinforcement may influence it. This kind of memory model inspired by research in cognitive psychology implies as well mechanisms for automatic "cognitive garbage collection".

During project's lifespan, the underlying OpenRobots ontology is expected to be collaboratively refined to better suit the requirements of the different partners involved in the project. To this end, the LAAS proposes to set up an on-line tool for collaborative development of the ontology.

4.7 Safety

BERT2 two will be equipped with torque and shell contact sensor in order to sense collisions with humans or objects. In addition, the low level joint controllers of BERT2 will be monitored and communication/bus error as well as hardware failure will be detected. BERT2 will provide a stream of health data for all joints (Physical Safety 1).

BRL will provide a higher level physical safety module (Physical Safety 2) which is able to evaluate reference and actual trajectories of any of the three robot platforms and will be capable of providing simple stop commands to the robot, given that the robotic platform features the corresponding interfaces.

Behavioural safety will be addressed by two distinct modules one forming part of the supervision system (BS1) and a further module residing between the high level motor command interface and the joint level trajectory generation (BS2).

5. Summary and Perspective

5.1 In response to the Technical Review Report 01/03/2009 to 28/02/2009

Because D5 takes place after the First year review (Period covered by the report: from 01/03/2009 to 28/02/2009), we take this opportunity to respond to that report with respect to the systems engineering component.

It may have been misunderstood that the project would only address functions that are defined in the SSE. This misunderstanding is likely due to our presentation of deliverables including D3. In the paragraphs below we will make it clear that as planned from the outset, the SSE represents only an example of the types of functions we will address. As identified in Section 3, there will be multiple additional scenarios defined, and implementation of the corresponding cooperative functions.

We refer to (A) the cover letter, and (B) the report itself.

(A) From the review cover letter:
Subject: Grant Agreement FP7 215805 – CHRIS "Cooperative Human Robot Interaction Systems"
Outcome of the First Periodic Review (Period 01/03/2008-28/02/2009)

R1 – An additional deliverable, presenting the motivations and the description of the additional scripted scenarios, has to be submitted to the Commission by M24.

(B) From the First period report

b. Overall recommendations

The reviewers noted two primary areas that require improvement. It is the opinion of the reviewers that both of these issues can be addressed by the team directly without modification to the current contract and without the requirement of additional immediate deliverables.
R1. First, to prevent the focus of the research from becoming too narrow, the consortium should describe additional scripted scenarios that will also become included in the long-term plans of the consortium.
R2. Second, the team needs to align the external documentation provided to the public and to the EU reviewers to match many of the decisions that have been made internally but not yet reflected in written form.

WP 3 – Systems Engineering: As part of WP3, the team delivered a set of functional requirements (D3) which outline the future direction of effort. The team departed from the original work specification by establishing a single Scripted Scenario Enactment (SSE) in place of a set of "Wizard of Oz" simulations. The reviewers noted that the entire team seemed to have input on this deliverable. While the reviewers agreed that focusing on a single scripted scenario provided a focus for describing the required modules that should be constructed in an initial robotic implementation, there was also a risk that focus on a single narrowly defined problem would result in a fragile system that generalized poorly. Discussions during the review meeting revealed that the team had many plans beyond this single SSE, but these were not contained within any of the available documentation. The team is encouraged to produce a detailed written specification to bring the externally visible aspects of the project. In doing so, the team should also make clear plans to provide documentation of the systems (perceptual and otherwise) which will not be part of their research efforts but that will be required for the success of the project.

Response: In reference to the statement in the review: The team departed from the original work specification by establishing a single Scripted Scenario Enactment (SSE) in place of a set of "Wizard of Oz" simulations.

We note that there is a slight misunderstanding, that is likely due to a lack of clarity in our presentation. In the Annex 1 we specified "*The consortium has already established that, as a priority, an agreed set of 'choreographed' bench-marked interactions must be set. Furthermore, we have agreed that these interactions manifest as the of 'Wizard-of-Oz' simulation in which the robot's interaction with a human is controlled (for example by a human actor or placing the robot limbs or even via limited tele-operation by a human or some combinations). This fixes the constraint envelope for the robot. These enactments specify the interactions and provide a set of comparative benchmarks when the robot has to interact with a human as an autonomous entity.*"

Thus, the Wizard-of-Oz simulation, which is now called the Scripted Scenario Enactment (SSE) was specified in Annex 1, and was developed as a starting point which would embody a set of behavioral capabilities to be implemented in the CHRIS project. What was unclear was the perception that this would be the only scenario that we would address. Indeed, we stated in Annex 1 (p 20)

> *Specification for the scenario testing of human robot interactive cooperation, performed safely within non-prescribed tasks, is performed in WP3 at the outset, following the initial meeting and simulation. In addition, information from WP6 regarding the measurement of cooperative action is incorporated in the test scenarios.*

Thus, according to the plan in Annex 1, we have undergone the analysis of the original SSE, and as illustrated in Section 3 of this document, we are developing new scenarios to extend the capabilities identified in the original SSE and in light of evidence from D4, Experimental data on children engagement in cooperative tasks.

5.2 Summary of Commercial Off the Shelf (COTS) vs. Original Contributions:

Perception: Cooperative human-robot interaction requires a baseline capability for perception and motor control. This baseline allows the development of cognitive capabilities for cooperative action planning and control, within an overall context of safety. Within the project, we will use existing technology for robot perception including vision segmentation and recognition capabilities, and non-vision based motion capture. *Given these existing perceptual capabilities we will contribute state of the art capabilities for action perception including the recognition of actions in the context of cooperative object manipulation and the attribution of agency. We will also contribute state of the art capabilities for perspective taking, and determination of perceptual and physical affordances available to the robot and the human based on analysis of the scene which has been reconstructed based on perception.*

Safety: Cooperative human-robot interaction crucially requires that the robot operates safely both with respect to its own safety, and with respect to safety for the human. *Within the project we will make contributions to the state of the art in both of these aspects.*

Cooperation: Finally, at the crux of Cooperative human-robot interaction is the notion of cooperation. Here the CHRIS project will make its most important contribution to the state of the art. Through the contributions of MPG via WP6, the project will identify the fundamental functional components of cooperation including the ability to take the perspective of the user, and more generally to develop a "bird's eye view" of a cooperative interaction in order to take the role of either participant (if physically possible) in a cooperative interaction.

This is made possible via close interaction with MPG and the associated deliverables from WP6. Deliverable **D4 Experimental data on children engagement in cooperative tasks** was successfully delivered at month 16. Already, based on this deliverable, and a continuous interaction between MPG and the other partners, these fundamental concepts of cooperation have been incorporated into both the original SSE, and the new scenarios that are identified in Section 3. Crucially, with Deliverable **D8 Conceptual framework of cooperation in human and HRI** at month 28, an implementable, framework for cooperation will be specified. Indeed, the translation from human experimental analysis of cooperation to technical implementation of the underlying capabilities is highly non-trivial, and is part of the fundamental research contribution of the CHRIS project.

Knowledge representation: the CHRIS project should enable also, along the same stream, the development of solid foundations for embodied cognitive systems in human-robot interaction context. This field is pretty new, and ontology-based approaches are especially new and promising. For instance, the ontology server might enable broader experiments regarding memory management and knowledge grounding.

Appendix 1: Snapshot of YARP ports

The Yarp framework provides port interfaces to essentially all of the functional capabilities in the system including the motor control level, scene perception, and components of state information that contribute to the knowledge base. Here we show a snapshot of a typical configuration of the existing YARP ports.

These are ports for our Neural Simulation Model neural network models to interact with the head:

/NSL/control/icub/head/command:o (tcp://10.0.0.10:10052)
/NSL/control/icub/head/rpc:o (tcp://10.0.0.10:10042)
/NSL/control/icub/head/state:i (tcp://10.0.0.10:10062)
/Port/In (tcp://10.0.0.10:10012)
/Port/Out (tcp://10.0.0.10:10002)

These are ports to “virtual devices” for recognition of actions and physical events

/actionRecognition/actions:o (tcp://10.0.0.14:10242)
/actionRecognition/events:i (tcp://10.0.0.14:10252)
/actionRecognition/events:o (tcp://10.0.0.14:10222)
/actionRecognition/objects:i (tcp://10.0.0.14:10232)
/actionRecognition/rpc (tcp://10.0.0.14:10262)

These are self explanatory

/icub/cam/left (tcp://10.0.0.2:10022)
/icub/cam/right (tcp://10.0.0.2:10012)

/icub/head/command:i (tcp://10.0.0.2:10082)
/icub/head/rpc:i (tcp://10.0.0.2:10182)
/icub/head/state:o (tcp://10.0.0.2:10072)

/icub/inertial (tcp://10.0.0.2:10152)

/icub/left_arm/command:i (tcp://10.0.0.2:10172)
/icub/left_arm/rpc:i (tcp://10.0.0.2:10062)
/icub/left_arm/state:o (tcp://10.0.0.2:10162)

/icub/left_leg/command:i (tcp://10.0.0.2:10232)
/icub/left_leg/rpc:i (tcp://10.0.0.2:10122)
/icub/left_leg/state:o (tcp://10.0.0.2:10222)

/icub/right_arm/command:i (tcp://10.0.0.2:10142)
/icub/right_arm/rpc:i (tcp://10.0.0.2:10242)

/icub/right_arm/state:o (tcp://10.0.0.2:10132)

/icub/right_leg/command:i (tcp://10.0.0.2:10202)
/icub/right_leg/rpc:i (tcp://10.0.0.2:10092)
/icub/right_leg/state:o (tcp://10.0.0.2:10192)

/icub/torso/command:i (tcp://10.0.0.2:10112)
/icub/torso/rpc:i (tcp://10.0.0.2:10212)
/icub/torso/state:o (tcp://10.0.0.2:10102)

/icub/view/left (tcp://10.0.0.1:10012)
/icub/view/right (tcp://10.0.0.1:10022)
/icubsrv (tcp://10.0.0.1:10002)
/pc104 (tcp://10.0.0.2:10002)
/root (tcp://10.0.0.1:10000)

This provides a triangulated stereo vision 3D (robot centered) reconstruction of all visible objects

/vision/3d/objects:o (tcp://10.0.0.14:10172)

Appendix 2:

D5.1 Adaptive Planning Capability Report (Internal Deliverable)

CHRIS

Cooperative Human Robot Interaction Systems

D5.1 Adaptive Planning Capability Report

Internal Deliverable

Produced by WP5: Rachid Alami, Emrah Akin Sisbot

Grant Agreement number: 215805

Project acronym: **CHRIS**

Project title: **Cooperative Human Robot Interaction Systems**

Project start date: 01 March 2008

Funding Scheme: **Small or medium-scale focused research project (STREP)**

Date of latest version of Annex I against which the assessment will be made: **14th Dec 2007**

Project coordinator name, title and organization:

Prof Chris Melhuish, Director, Bristol Robotic Laboratory (BRL), University of the West of England, Bristol

Tel: ++44(0)1173286334

Fax: ++44(0)1173282688

E-mail: Chris.Melhuish@brl.ac.uk

Project website address: http://www.chrisfp7.eu/

Version: Final, Rev.1
Due date: October 30, 2008
Delivery date: Draft (Jan. 23, 2009), Final (Jan. 30, 2009), Revised (Feb. 15, 2009)

Contents

1 Executive Summary

This document deals with the definition of the planning capability of the desired CHRIS robot.

It is based on a preliminary work that allowed to identify the main subjects that will be investigated further in the project.

We identify the main components linked to planning, and more generally, decisional capabilities through an analysis of the CHRIS scenarios.

We then discuss the main topics that will be investigated in order to elaborate models and develop protocols and algorithms that should endow the CHRIS robot with human-friendly collaborative abilities.

Finally, we briefly present a set of software environments and experimental tools that we will develop and deploy to support the research activities for WP5 in close interaction with the other project workpackages.

2 Introduction

The CHRIS context of Human-Robot Interaction (HRI), that has been discussed and refined collectively in the first months on the project [1], involves a robot that shares space with a person and that is able to cooperate by participating to a manipulation task[2].

This has a great influence in terms of the main aspects to be dealt with at the decisional level. Indeed, it raises several very interesting and challenging issues:

- the ability of the robot to conduct intricate human-robot cooperation in a common space and the need to do this in a natural manner with respect to the person,
- the importance of contingency and uncertainties (from the robot point of view) due to the presence of the person,
- the interplay of symbolic reasoning and geometrical reasoning.

We have conducted investigations for the design of a conceptual architecture of the decisional component of a robot companion that acts, learns and interacts with humans.

This architecture is aimed to be a framework that provides a basis for a principled way to deal with robot task achievement in presence of humans or in synergy with humans.

Section 3 discusses the CHRIS context and the main challenges relative to adaptive planning. Section 4 proposes a decomposition of the planning abilities into components that are examined in the sequel. We then present the tools that will be developed in a first step.

3 The CHRIS context

In a context where a robot and a person share a joint collaborative task, the robot reasoning mechanism should not only involve task and action management but should also include spatial reasoning on the environment, on the humans, on the objects, and on the robot itself. As the robot may be in close interaction with its human partner, it should not only ensure the feasibility of its part of the joint task but also should satisfy safety and legibility requirements of its motions.

In order to equip the robot with such behavior, a sophisticated reasoning mechanism explicitly designed for Human-Robot Collaborative Interaction is needed.

3.1 Approach

From the general backgrounds of robotics, we can identify three activities involved in the robot control, in addition to the sensory-motor functions that implement sensing, locomotion or manipulation. Those are three strongly interdependent activities:

- reactive planning
- knowledge-based reasoning and maintenance of robot knowledge
- deliberative planning

The objective of WP5 is to study and contribute to these issues through the perspective of human robot interaction. Indeed, they correspond, globally, to the three research streams that have been defined:

[1]Meetings and Internal workshops at Bristol, Genoa and Toulouse, Visit of LAAS to MPG.
[2]See also D3 "Systems Engineering Specification"

- T5.1 Engagement management
- T5.2 Uncertainty management
- T5.3 Goal and Decision Planning

3.2 Scenario

We place ourselves in a scenario where a robot and a person are sitting face to face around a table (Figure 1). The robot, being the HRP-2 humanoid platform of LAAS, has dual-arm (7 dof each) manipulation capability. A number of various sized objects are placed on the table, some only reachable or visible to the robot, some only to the human. In this scenario the task mainly involves, physically,"pick & place" and "pick & transfer" actions.

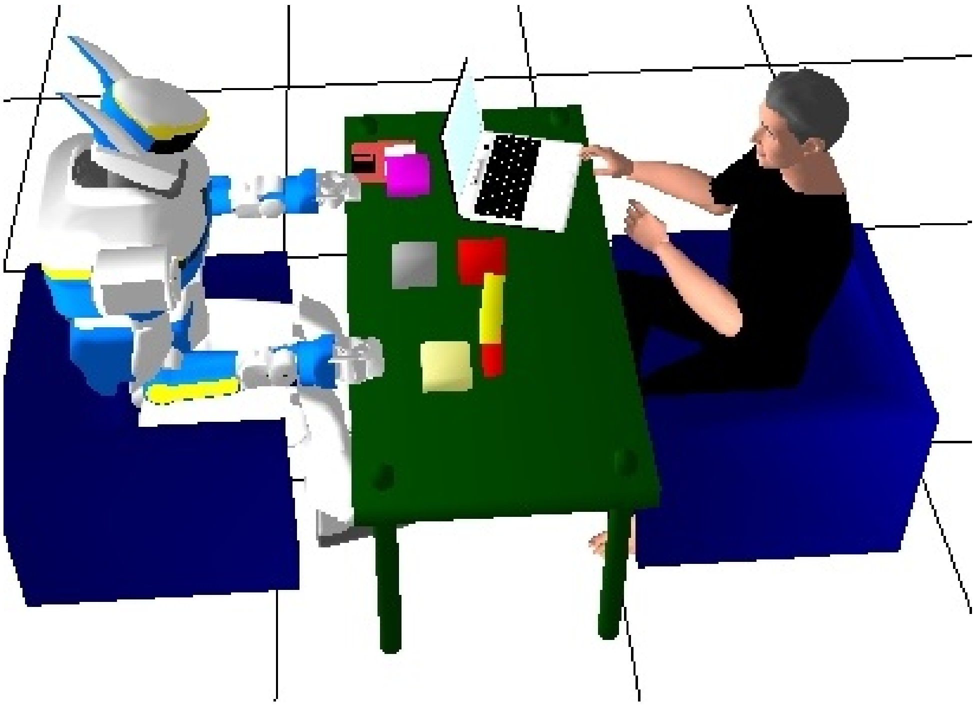


Figure 1: A scenario where a robot (HRP-2) and a person are sitting face to face to accomplish a collaborative task.

3.3 Challenges and Methodology

The human presence brings new requirements for robot's abilities both at the functional and at the deliberative levels[Klein 04]. For the former, the topics involve motion[Kulic 07, Berg 04, Madhava Krishna 06], navigation[Althaus 04, Sisbot 07b], manipulation[Kemp 07] in presence of humans as well as perception of human activities[Breazeal 01, Burger 08]. For the latter, when interacting with humans, robot needs to incorporate communication and collaboration abilities.

We examine here below the deliberation issues linked to both aspects.

3.3.1 Deliberation for cooperation

Several theories dealing with collaboration [Cohen 91, Bratman 92, Grosz 96, Clark 96] emphasize that collaborative tasks have specific requirements compared to individual ones e.g. since the robot and the person share a common goal, they have to agree on the manner to realize it and they must show their commitment to the goal during execution. Several robotic systems have already been built based on these theories [Rich 97, Breazeal 99, Sidner 05, Tambe 97, Breazeal 03] and they all have shown benefits of this approach. They have also shown how difficult it is to manage turn-taking between communication partners and to interleave task realization and communication in a generic way. Finally, today only few systems [Fong 06, Breazeal 03, add Aurelie Clodic 08] take humans into account at all levels.

3.3.2 Cooperative motion and manipulation

A robot that will serve as a helper among humans should not only be a machine but it should respect social rules and protocols [Chatila 02][Fong 03] ensuring a:

- *Safe motion*, i.e., that does not harm the human,
- *Socially acceptable motion*, i.e., that takes into account the comfort of the human as well as his preferences and needs[3],
- *Reliable and effective motion*, i.e., that achieves the task adequately considering the motion capacities of the robot.

The robot has to perform actions while determining where a given task should be achieved, how to place itself relatively to a human, how to approach him/her, how to hand the object and how to move in a relatively constrained environment in the presence of humans.

All these actions should be accomplished without requiring extensive amount of human effort towards the goal of facilitating humans' tasks. In order to avoid a "human-centric" interaction, the robot should reason by its own on its human partner's capabilities, his accessibility, his image of world, and should have a complete understanding of the task and the environment.

We define three important research directions each being a geometric reasoning and planning skill for a cooperative robot:

- Perspective Taking,
- Spatial language for interactive manipulation,
- Human-Aware motion.

4 Architecture of the robot decisional activities

We have devised a control architecture dedicated to robot decision and action in a human context (Figure 2). It has been developed as an instance of the generic LAAS Architecture for Autonomous Systems [Alami 98a].

The decisional layer can be organized into three interdependent components:

- A Control component that is essentially reactive since it has to comply with real-time constraints. In the LAAS architecture, it is called decisional kernel or Supervisor. It is based on an incremental context-based task refinement in a human context.

[3] When the intention of the robot is clear (legible) then it adds also to safety.

- The Knowledge Base; which embeds all the knowledge available to the other components: models of the environment, of the task, of the robot and its abilities as well as, and this is a key issue here, models of the human partner, his abilities and preferences (from the point of view of the robot).
- The Planner. We will see below that it covers several aspects. In the framework of CHRIS, we propose to call it "Human Aware Planners" (HAP) since the robot has to take into account explicitly the presence and or the activity and intentions of its human partner.

These components are aimed at providing integrated abilities to support human-robot collaborative task achievement as well as capacities to elaborate task plans involving humans and robots and to produce legible and socially acceptable behavior. We analyse them in the next three sections.

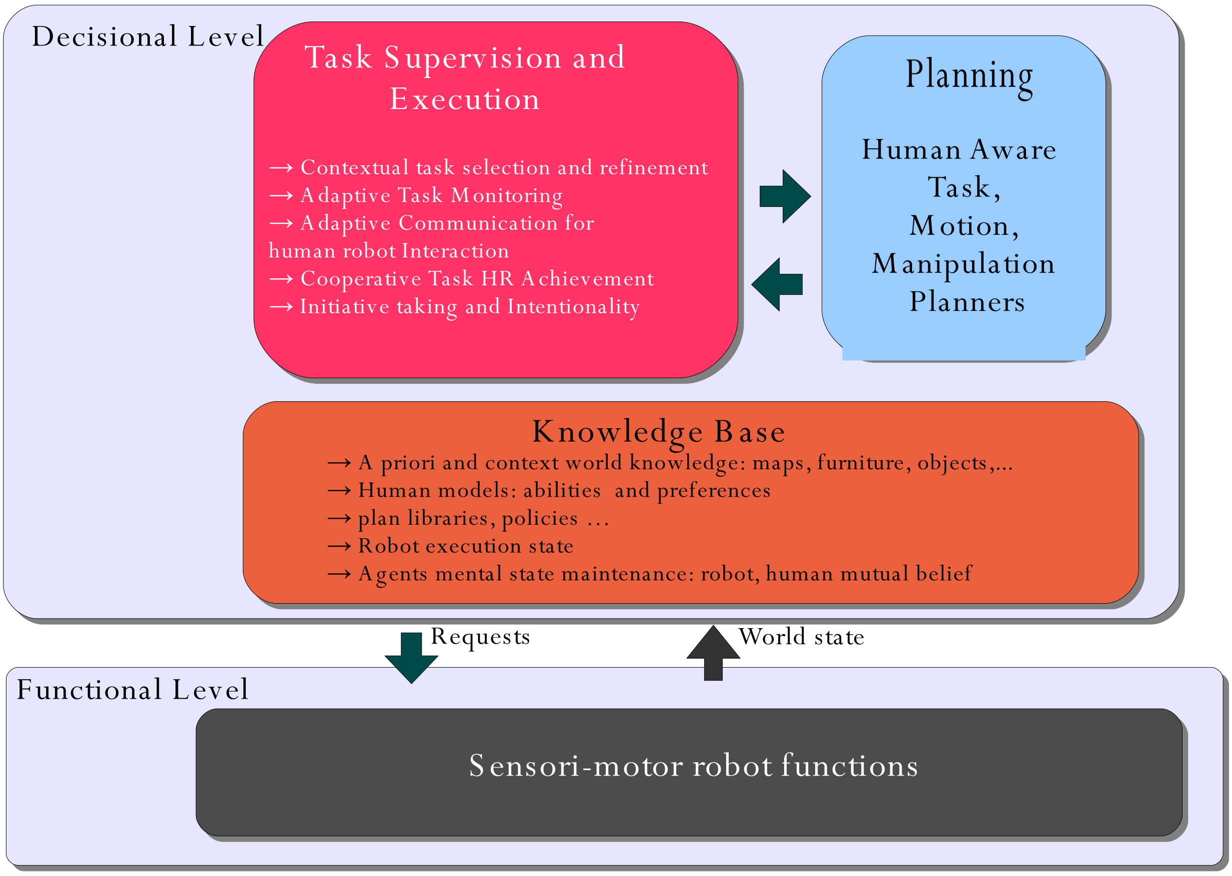


Figure 2: The proposed control architecture for interactive robots

5 The Decisional Kernel

Designing a supervision system for an interactive robot raises several new challenges.

In fact, when performing tasks in interaction with humans, the robot supervision system is not only responsible for the refinement and the correct execution of the robot plan, but also for the appropriate set of communications and monitoring activities within and around task realization. It is also in charge of monitoring human commitment and activities in order to provide appropriate response based on the current context (which is included in the fact database). Furthermore, robotics context implies that execution can fail. Since the robot is evolving in the real world, hardware and software failure may occur frequently. To ensure human safety, the supervision system must be capable of recovering from failure by stopping current activities and re-planning task execution.

Another ability in the CHRIS context would be initiative taking, i.e. mechanism that permits the robot to exhibit a proactive behaviour, for instance taking the initiative to serve a drink or to behave as a "curious" robot that decides to acquire information about the state of the environment (e.g. exploration of new objects placed by a person on a table).

LAAS is currently working on such a HRI-enabled supervision system, called SHARY[4]. It can be surely used as a basis for experimenting.

The originality of SHARY, as a supervision system, lies in its ability to take into account not only the task achievement but also communication and monitoring needed to support interactive task achievement in a flexible way. SHARY allows to define a task or a hierarchy of tasks linked to more or less elaborated "communication policies"[Clodic 07b] that enable to execute tasks and the possibility to deal with contingencies that could occur during its execution (or even to stop the task in case of unexpected events).

We have defined a set of communication acts that we found mandatory in the framework of task achievement [Clodic 07a]. At any time, both the user and the robot can propose the following task-based acts:

- `ASK-TASK`: Proposing a task.
- `PROPOSE-PLAN`: Proposing a plan (recipe) for a given task.
- `MODIFY-PLAN`: Proposing a modification of the current plan for a given task.
- `GIVE-UP`: Gives up a task (e.g. because the task becomes impossible). For the robot this is a way to announce that it is unable to achieve the task.
- `CANCEL`: Cancellation of a task (voluntary give-up).
- `END-TASK`: Announces that the task has been done.
- `REALIZE-TASK`: Announces that the task performance will start.

This set takes inspiration from Joint Intention Theory ([Cohen 91]) that states that each partner should be informed of the beginning, realization and ending of a joint task.

5.1 Investigating cooperative skills

As mentioned above, a main point would be to refine, through multi-disciplinary work, the categorisation, the models and the articulation between the different abilities that should be available at the robot level.

This will be based on results obtained by MPG and by their studies on human, on other species and on their comparison.

First, there is an interesting identification and categorisation of the abilities involved in shared activity. Some can be considered as individualistic (in CHRIS, robot-centric) even if

[4]SHARY: Supervision for Human Aware Robot Ynteraction

they involve perception or action in presence of other individuals. [Tomasello 07] mentions, for instance, the need for powerful intention reading and cultural learning.

In the case of CHRIS, the robot intention reading and "cultural learning" issues could be partially fulfilled by cognitive abilities in terms of human and robot manipulation activities (see §6.1).

But, and this is essential here, [Tomasello 07] explains that there is crucial difference between human cognition and that of other species. It is the ability to participate with others in collaborative activities with shared goals and intentions: shared intentionality.

We will work into refining this notions and investigate ways to endow the robot with this ability, i.e. define principled ways to build in the robot the capacity to share psychological states with others and unique forms of cognitive representation for doing so.

Other questions are linked, more generally to the underlying motivations and how they could be linked to the construction, at the level of the robot of initiative taking.

At the level of the robot decisional kernel, we will study, based on this cooperative work, how the robot can share a task with a human partner, in which conditions a goal appears in the "system" and how the robot "negotiates" the tasks to achieve it.

Of interest are also questions linked to multimodal dialog (including gestures) that supports these activities.

5.2 Integration of learning

One desired feature in the targeted system, as discussed in CHRIS "Description of Work" document and illustrated in Deliverable D3, is the integration of learning. This is essentially sequence or script learning. It is different from sensory-motor skill learning.

Consequently, this ability is clearly situated at the level of the control program.

An interesting approach is the one proposed by [Kirsch 08] where a so-called Robot Learning Language has been defined which makes learning tasks executable within the control program. Upon the completion of the learning process, the learned function is integrated in the control program.

Another approach that we would like to investigate is to treat a learning process as a human-robot cooperative problem solving process based, for instance on [Lochbaum 98, Pollack 03, Montreuil 07]

6 The Knowledge Base

As already mentioned above the Knowledge Base embeds and maintains all the knowledge available to the other components: models of the environment, of the task, of the robot and its abilities as well as, and this is a key issue here, models of the human partner, his abilities and preferences (from the point of view of the robot).

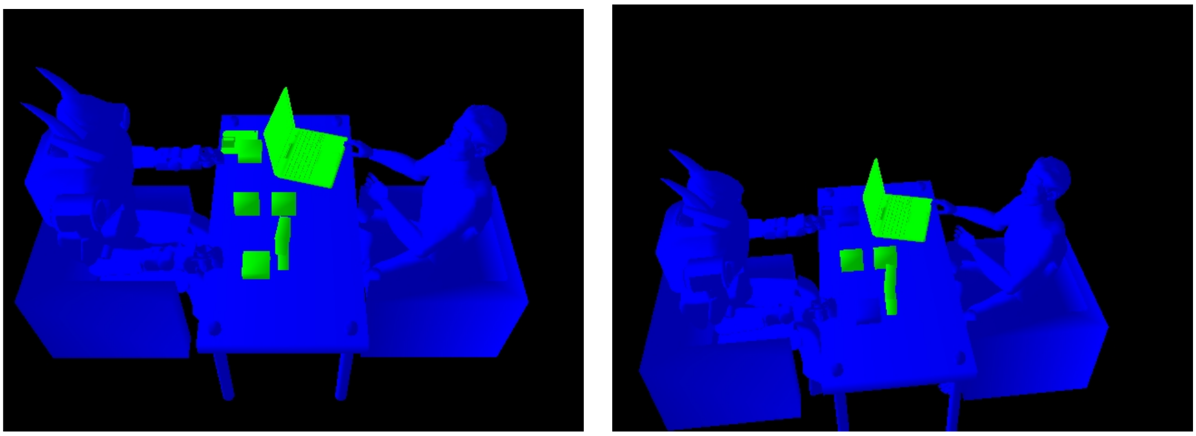
It is not only a database but an active system which maintains through elaborate reasoning and inference mechanisms a coherent view of the robot knowledge.

Of interest in the context of CHRIS will be the models of the human, the mental state of the robot and the management of the Human-Robot mutual beliefs.

6.1 Perspective Taking

The notion of "perspective taking" comes from psychological studies on human to human interactions [Tversky 99, Lee 01]. It refers essentially to the fact of reasoning from other person's point of view. It is also interpreted as taking its own perspective from a different place on the space by applying "mental rotation" [Ackerman 96] to perceive the environment. These sets of actions are used by humans in their everyday lives, and are intended to alleviate communication between individuals and to help to have more efficient interactions.

In a HRI scenario, a social robot is expected to possess perspective taking capabilities to reason on the human partner's perspective in order to infer his capabilities and needs for task's realization. An important step towards human-like perspective taking is "geometric reasoning on perspective taking", meaning that the robot will put itself mentally at the human's place to reason on his "perception".



(a) Object perceived by the robot. (b) Object perceived jointly by the robot and the human.

Figure 3: Perception differences between the robot and its human partner.

Figure 3 illustrates the perception difference between the robot and the person in figure 1 scenario. This difference clearly indicates that there are some objects on the table which are invisible to the human but visible to the robot. In such a situation, the robot needs to put itself mentally to its human partner's perspective, to reason on what he perceives and to act in order to help him (fetch something invisible or too far to be reached easily by the person). For example, in figure 1, if the human requires the orange disc, which is invisible to him, the robot should foresee that the human will need it and that it is invisible to him. So the robot should infer that it needs to either point the disk, or pick it and hand it to the person.

PersPective Placement (PSP) mechanism [Marin 08] developed at LAAS-CNRS is a first approach towards this problem. PSP reasons on what human perceives by allowing the robot to put itself in human's perspective. This system allows the robot to evaluate and generate most suitable robot configurations to human's perception. Although PSP can reason on people's perception, the robot should also reason on human's kinematics and his capabilities at a given moment in order to infer the help the person will need to accomplish his task.

The robot should maintain a knowledge base on human containing which object does the human see and which he does not and which objects he can reach/manipulate and which he cannot.

As the robot's image of the world depends highly on its sensors, the errors of perception on objects as well as human's positions may cause an incoherence between the real environment and robot's image. Moreover the objects can be moved without the robot noticing (e.g moved by the human partner). In order to take into account these facts, the objects in the environment, including the human's kinematic structure, should be modeled with an attached uncertainty. This uncertainty will increase unless the robot perceives the object again.

6.2 A Basis for a Spatial Language for Interactive Manipulation

Another aspect that one should consider is the richness of a physical Human-Robot Interaction. Because of the unpredictable nature of its human partner, the robot should be ready (equipped with sufficient cognitive tools) to understand and act according to human's requests.

As all the actions cannot be modeled beforehand, basic robot actions will be defined and will be represented as verbs in a geometrical language. Depending on robot's physical structure, its basic actions should include:

Grasp (something) : Robot reaches towards an object, a person or itself and grasps it.

Give (something) to (someone) : Robot hands over what it has its hand to specified person.

Touch (something) : Robot reaches specified thing and touches it.

Push (something) : Robot reaches and pushes specified object or person.

Look at (something) : Robot look at specified thing.

Show (someone) (something) : Robot shows or points asked thing to a person.

Place (something) on (somewhere) : Robot places the object that it possesses onto the specified surface.

Hold (something) : Robot holds an object still.

With this simple action vocabulary and their attached functionalities, the human can describe his requests more efficiently without further need for a more precise information. By combining these actions with "objects" and "humans", the robot can build simple tasks. For example, in response to human's request "Show me the orange disc" the robot can infer that this is a "Show" action with "Orange Disc" and "Me (the human)" as parameters. By reasoning on human's perspective, it can generate a posture where its hand is sufficiently visible to the human and is pointing towards the object.

Another important property that comes to surface with the usage of such language is the properties that objects acquire from their spatial placement. An object as well as the human and the robot have a number of properties that came from their existence, for example, a table has the property to contain objects on top, a human has the property to grasp an object. In addition to these natural properties, objects can obtain spatial properties depending on their place and on robot's perception. For the robot an object can have the properties "Visible", "Next to another Object", "Allow manipulation", etc. With these properties the robot can have symbolic beliefs on geometric properties of its environment.

6.3 An ontology

We will investigate issues of grounding the internal information used by the robot for its actions and perceptions to abstract concepts that can be dealt with in an ontology that can be used to describe human and robot objects manipulation.

One very interesting contribution, for instance, is given in [Tenorth 08]. This could allow more natural interaction and even build an understanding of underspecified action commands by the human.

We will investigate how advanced models of the manipulation planning problem as expressed in [Gravot 03, Cambon 04] serve as a basis for a principled method of grounding object manipulation problems by robots and humans. Besides, geometric computation of perspectives and more generally of affordances linked to the language defined above can be also used for grounding.

7 Human-Aware Planning

We consider here the different planning capabilities: symbolic task planners and geometric motion planners. As mentioned above, they can be characterized, in the framework of CHRIS, by the necessity to take into account explicitly the human.

7.1 Symbolic task planning

The problem here consists in elaborating plans for a robot that will allow it to cooperate with human. Criteria will be the utility and the fact that the robot behaviour based on the synthesised plans is "socially acceptable".

An instance of such a system is a planner, called HATP (for Human Aware Task Planner) that we are currently developing.

It is specially designed for the interactive action between heterogeneous agents, in our case humans and robots. It is based on hierarchical task planning[Ghallab 04] which gives to the planner some ability to answer to certain issues defined above. It also integrates "social behavior rules", which orient the robot decision toward the synthesis of socially acceptable plans.

We define a social rule as a convention which associates penalties to a behaviour of an agent in a given context. In human environments most of these rules are implicit between the members of a community and they vary from a society or a culture to another. In HATP we are investigating rules in the domain description language of symbolic task planners. Examples of such rules involve

Undesirable states : i.e world states which are dangerous, unpleasant or inadequate. Example: the robot lets the fridge door open, the robot puts an object on the floor which may impede the movement of other agents or, more dangerous, leaves the gas turned on.

Undesirable sequences : Some action combinations can conduct to an uncomfortable feeling for human partners. For example, a plan in which the robot puts down an object and its partner picks it up immediately after might be considered as an awkward sequence if a the robot can hand over directly the object.

Bad decompositions : This rule is used to select better choices at a high level of decomposition. The idea is that specific decompositions must remain possible but they must be used only if necessary. For example, when the robot has to leave an object for someone, it is better to place it on a "visible" place.

Effort balancing : The idea is to establish a balance of effort among partners. If we have a team of two robot agents, the amount of effort could be the same for all the staff. If the team is composed of a human and a robot, then the robot must provide more effort than the human.

Intricate links between plan streams : Indeed, unless necessary, it is clear that multiple agent plans that are synchronized by frequent synchronization links should be avoided. Such links make plans fragile and introduce a lot of dependency between agents.

We intend to develop and integrate such a planner. Several issues here are of interest: the consideration of uncertainty on one side and the introduction of temporal reasoning on the other[Pollack 03].

7.2 Human-Aware Motion Planning

Robots in close interaction with humans bring new questions and concerns about safety of the humans sharing the same environment with the robot. These concerns leads the notion of "safety" to be studied in detail (e.g. [Alami 06]) and to be evaluated through user studies (e.g. [Haddadin 08]). Nonaka et al[Nonaka 04] studied the concept of safety by two aspects: "physical" safety and "mental" safety of human. Physical safety means that the robot does 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.

These studies converge towards a common classification of safety strategies originally introduced in [Ikuta 03]. In this classification, safety strategies are divided into two main categories: design strategies, where safety is ensured by the mechanical hardware design of the robot; and control strategies, where safety is ensured by controlling the motion of the robot. In the decisional level, we are interested in developing planning and control strategies to generate safe and easily understandable robot motions.

Research towards safe physical interaction in decisional level is growing rapidly because of the critical nature of the issue. Among various methods danger index minimization[Nokata 02, Ikuta 03, Traver 00, Kulić 05] is the most addressed because of its reactivity. Yet, this method remains as an execution solution rather than a planning approach and ensures only the physical safety of the human.

LAAS has developed a Human-Aware navigation[Sisbot 07b] and manipulation[Sisbot 07a] planning framework that generates robot paths not only physically safe but also comfortable to the human. This framework explicitly takes into account the human by reasoning on his partial kinematic structure, his accessibility, his field of view, his posture, and his activity. Two motion planners are resulted from this work and both integrated[add Aurelie Clodic 08] to real robotics platforms.

Being a novel approach, this work opened many research questions some of which will be studied in CHRIS Project:

- Reasoning on human's full kinematic structure: Human's full physical structure needs to be taken into account in order to produce safer motions. In scenarios where the robot and the human are in close interaction, this notion becomes more crucial as every part of the person in the threat range of robot motions.

- Humanoid robot integration: Extension of roadmap based dual-arm manipulation planning schemes[Gharbi 08] into this framework will open challenges of adapting probabilistic approaches to real world execution. This extension will later be integrated to HRP-2 humanoid robot.

- Taking into account uncertainty: Human's unpredictable nature and robot sensors' detection errors avoid the precise execution of planned motions. In order to adapt to the unexpected, uncertainty on human's and objects' positions should be taken into account in the planning level. In this frame, existing methods[Alterovitz 07, Burns 07] will be investigated.

8 Software environment and experimental tools

We intend to develop and deploy a number of software systems and tools as a basis for experiments or as environment that will allow the development of the desired algorithms.

8.1 Robot Supervision

The current version of SHARY is essentially coded in OpenPRS [Ingrand 96]. OpenPRS is an open source version of PRS (Procedural Reasoning Systems) / Propice[5]

It is based on C-PRS and Propice, which were themselves inspired from the Lisp PRS originally developed at SRI International. PRS has been used for a wide variety of applications, from Mobile robot execution control to Space Shuttle operational procedure execution.

The programming paradigm is based on a BDI approach and is very convenient to encode goal-driven as well situation-driven behaviours.

We plan to provide an interface library using YARP[6] which should allow to facilitate its integration in the framework of CHRIS particularly to access the robots sensory-motor functions.

8.1.1 Ontology management

A software tool that will allow to build and exploit an ontology that will be used in human and robot objects manipulation.

8.1.2 Motion Capture

In order to study perspective taking and motion coordination between the robot and the human, we intend to integrate a motion capture system.

The motion capture system installed in LAAS has the ability to model and track kinematic structures in real-time. The human equipped with markers can be tracked easily with this system. In the scope of the project we plan to track the upper body of the human (i.e. both arms, torso, and the head) and the objects in robots environment.

A software environment will be developed in order to feed directly, in real time, the MOVE3D motion planning system trough a Genom/YARP protocol.

8.1.3 Motion planning and geometric reasoning environment

The future motion planner and the perspective placement mechanism will be developed and integrated into Move3D [Siméon 01] platform in C programming language. Both systems share the same models and communicate in the same software environment.

Move3D will be used to model, at the geometric level, the environment, the robot, the objects, human's kinematic structure as well as robot's sensors.

The whole system is encapsulated into a Genom [Fleury 97] module that enables the easy communication with other components and modules of the robot in LAAS Architecture [Alami 98b].

8.1.4 Human Simulation Environment

One of the most important practical challenges of HRI field is the necessity to include human in the development process. Users are permanently needed to design, develop, integrate, and test stages of the research. When naive users are present, safety measures need to be taken carefully with sufficient staff.

In order to avoid mobilization of many people and speed up the design and test processes, a human simulation environment will be developed. By placing and controlling virtual humans in a 3D virtual environment, the user will see on his screen what the virtual human is actually seeing (as shown in figure 4). The user will have control of basic actions of the virtual mannequin,

[5] see "http://softs.laas.fr/openrobots/wiki"

[6] see http://eris.liralab.it/yarp/

like move, turn, sit, look, and fetch. The available motions will be developed in parallel to the robot's spatial language (§ 6.2).

To see the environment in the robot's point of view, the simulator may also have an additional option to show what the robot sees with its cameras (as illustrated in figure 4).

The simulator will also enable simple human commands, e.g. "give me, show me, etc" and dialogue e.g. "hello, yes, no, ok, etc".

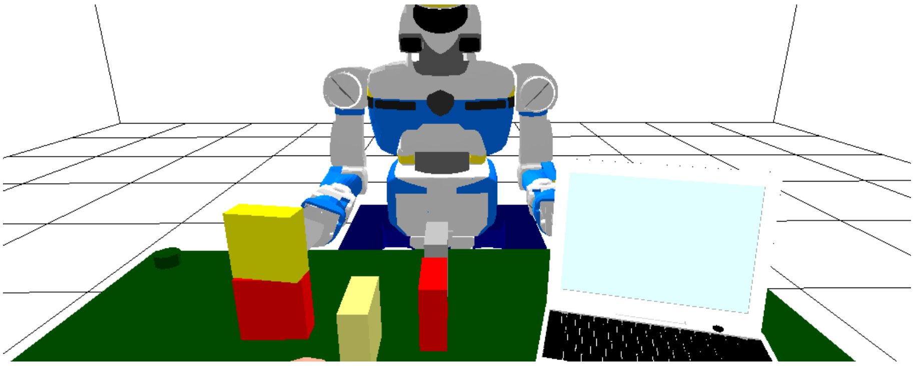


Figure 4: A window showing what human sees at a moment (figure1).

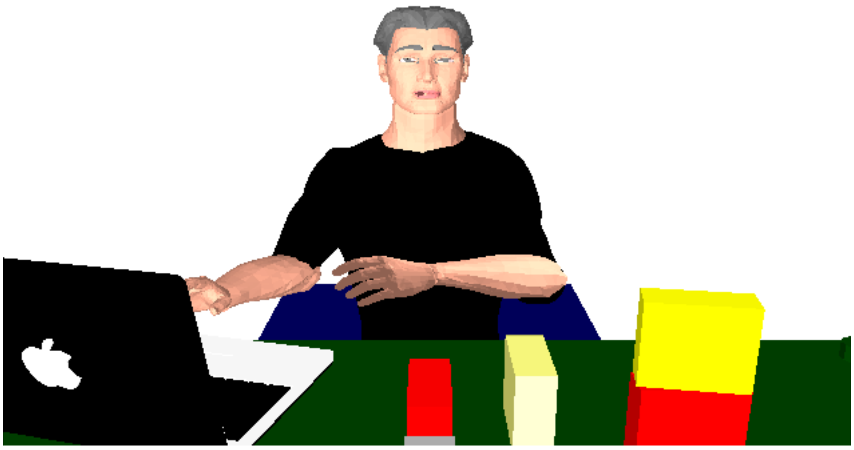


Figure 5: A window showing what robot perceives at the same time in the same scenario.

9 Next steps

We will pursue our investigation and refinement of the conceptual aspects of human-robot shared intentionality and activity.

In parallel, we intend to develop an interactive task planning and execution system that integrates models and guidelines provided by this first activity.

We will also develop and deploy a set of tools that will allow to experiment and illustrate the contributions relative to Decision Planning and its integration with the other topics developed in the project.

These activities will be conducted in the framework on three task:

- **T5.1 Engagement management:**

  The next activities will involve the study of shared intentionality and the incremental elaboration of models and the investigation on how they could conduct to an effective implementation.

  This activity will be conducted in relation with the refinement of the scenarios envisaged in CHRIS (see Annex 1)

- **T5.2 Uncertainty management**

  The work here will essentially involve:

  - The implementation of a prototype ontology in relation with the knowledge base.
  - The extension and implementation of geometric algorithms for Perspective-taking and computation of affordances
  - The implementation and integration of a first on-line estimation of Human and Robot affordances (PSP, Human Motion Tracking, Object identification and localisation)

- **T5.3 Goal and Decision Planning**

  The work here will essentially involve:

  - The implementation and tests of a prototype version of a task planner for cooperative human-robot task achievement (from joint goal to shared plans)
  - The study of planning algorithms for robot actions of the scenario (H-R cooperative object manipulation)

References

[Ackerman 96] Edith K. Ackerman. *Perspective-Taking and Object Construction: Two keys to learning.* In Constructionism in Practice: Designing, Thinking, and Learning in a Digital World, Mahwah, New Yersey, 1996.

[add Aurelie Clodic 08] Emrah Akin Sisbot add Aurelie Clodic, Rachid Alami and Maxime Ransan. *Supervision and Motion Planning for a Mobile Manipulator Interacting with Humans.* In ACM/IEEE International Conference on Human-Robot Interaction, HRI, Amsterdam, The Netherlands, 2008.

[Alami 98a] R. Alami, R. Chatila, S. Fleury, M. Ghallab and F. Ingrand. *An Architecture for Autonomy.* Internatonal Journal of robotics Research, Special Issue on Integrated Architectures for Robot Control and Programming, vol. 17, no. 4, 1998.

[Alami 98b] R. Alami, R. Chatila, S. Fleury, M. Ghallab and F. Ingrand. *An Architecture for Autonomy.* International Journal of Robotics Research, vol. 17, pages 315–337, 1998.

[Alami 06] R. Alami, A. Albu-Schaeffer, A. Bicchi, R. Bischoff, R. Chatila, A. De Luca, A. De Santis, G. Giralt, J. Guiochet, G. Hirzinger, F. Ingrand, V. Lippiello, R. Mattone, D. Powell, S. Sen, B. Siciliano, G. Tonietti and L. Villani. *Safe and Dependable Physical Human-Robot Interaction in Anthropic Domains: State of the Art and Challenges.* In A. Bicchi and A. De Luca, editeurs, Procceedings IROS Workshop on pHRI - Physical Human-Robot Interaction in Anthropic Domains, Beijing, China, October 2006.

[Alterovitz 07] Ron Alterovitz, Thierry Simeon and Ken Goldberg. *The Stochastic Motion Roadmap: A Sampling Framework for Planning with Markov Motion Uncertainty.* In Proc. Robotics: Science and Systems, June 2007.

[Althaus 04] 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.

[Berg 04] Jur P. van den Berg and M. Overmars. *Roadmap-based Motion Planning in Dynamic Environments.* Rapport technique, Utrecht University, NL, 2004.

[Bratman 92] Michael E. Bratman. *Shared Cooperative Activity.* The Philosophical Review, vol. 101, no. 2, pages 327–341, 1992.

[Breazeal 99] Cynthia Breazeal and Brian Scassellati. *How to build robots that make friends and influence people.* In Proc. IROS99, Kyonjiu, Korea, pages 858–863, 1999.

[Breazeal 01] C. Breazeal, A. Edsinger, P. Fitzpatrick and B. Scassellati. *Active vision for sociable robots.* IEEE Transactions on Systems, Man and Cybernetics, Part A, vol. 31, no. 5, pages 443–453, September 2001.

[Breazeal 03] C. Breazeal. *Towards Sociable Robots.* Robotics and Autonomous Systems, pages 167–175, 2003.

[Burger 08] B. Burger, I. Ferrane and F. Lerasle. *Multimodal Interaction Abilities for a Robot Companion.* In ICVS, pages 549–558, 2008.

[Burns 07] B. Burns and O. Brock. *Samplingbased motion planning with sensing uncertainty.* In IEEE International Conference on Robotics and Automation, 2007.

[Cambon 04] S. Cambon, F. GRAVOT and R. ALAMI. *A robot task planner that merges symbolic and geometric reasoning.* In 16th European Conference on Artificial Intelligence (ECAI'2004), pages 895–899, Valence (Espagne), 2004.

[Chatila 02] R. Chatila, R. Alami, T. Simeon, J. Pettre and L. Jaillet. *Safe, reliable and friendly interaction between humans and humanoids.* In IARP International Workshop on Humanoid and Human Friendly Robotics, pages 83–87, Tsukuba, Japan, December 2002.

[Clark 96] H. H. Clark. Using language. Cambridge University Press, 1996.

[Clodic 07a] A. Clodic, R. Alami, V. Montreuil, S. Li, B. Wrede and A. Swadzba. *A study of interaction between dialog and decision for human-robot collaborative task achievement.* In 16th IEEE International Symposium on Robot and Human interactive Communication, RO-MAN 2007, 2007.

[Clodic 07b] A. Clodic, M. Ransan, R. Alami and V. Montreuil. *A management of mutual belief for human-robot interaction.* In IEEE Int. Conf. on Systems, Man and Cybernetics (SMC), 2007.

[Cohen 91] P. R. Cohen and H. J. Levesque. *Teamwork.* Nous, vol. 25(4), pages 487–512, 1991.

[Fleury 97] S. Fleury, M. Herrb and R. Chatila. *Genom: a tool for the specification and the implementation of operating modules in a distributed robot architecture.* In IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, pages 842–848, Grenoble, France, September 1997.

[Fong 03] T. Fong, I. Nourbakhsh and K. Dautenhahn. *A Survey of Socially Interactive Robots.* Robotics and Autonomous Systems, vol. 42, no. 3-4, 2003.

[Fong 06] T. W. Fong, C. Kunz, L. Hiatt and M. Bugajska. *The Human-Robot Interaction Operating System.* In Human-Robot Interaction Conference. ACM, March 2006.

[Ghallab 04] M. Ghallab, D. Nau and P. Traverso. Automated planning - theory and practice. Morgan Kaufmann Publishers, 2004.

[Gharbi 08] Mokhtar Gharbi, Juan Cortés and Thierry Siméon. *A sampling-based path planner for dual-arm manipulation.* In IEEE/ASME International Conference on Advanced Intelligent Mechatronics, july 2008.

[Gravot 03] F. Gravot, S. Cambon and R. Alami. *aSyMov: a planner that deals with intricate symbolic and geometric problems.* In Invited paper. 11th International Symposium on Robotics Research (ISRR'2003), Sienne (Italie), 2003. Robotics Research, Eds. P.Dario, R.Chatila, Springer tracts in advanced robotics 15, Springer, ISBN 3-540-23214-1, 2005, pp100-109.

[Grosz 96] B. J. Grosz and S. Kraus. *Collaborative Plans for Complex Group Action.* Artificial Intelligence, vol. 86, pages 269–358, 1996.

[Haddadin 08] Sami Haddadin, Alin Albu-Schaffer and Gerd Hirzinger. *The role of the robot mass and velocity in physical human-robot interaction - Part I: Non-constrained blunt impacts.* In IEEE International Conference on Robotics and Automation, ICRA, pages 1331–1338, Pasadena, USA, May 2008.

[Ikuta 03] Koji Ikuta, Hideki Ishii and Makoto Nokata. *Safety Evaluation Method of Design and Control for Human-Care Robots.* International Journal of Robotics Research, vol. 22, no. 5, pages 281–297, 2003.

[Ingrand 96] F. Ingrand, R. CHATILA, R. ALAMI and F. ROBERT. *PRS: a high level supervision and control language for autonomous mobile robots.* In 1996 IEEE International Conference on Robotics and Automation (ICRA'96), pages 43–49, Minneapolis (USA), 1996.

[Kemp 07] C.C. Kemp and E. Edsinger A. Torres-Jara. *Challenges for robot manipulation in human environments.* Robotics & Automation Magazine, IEEE, 2007.

[Kirsch 08] A. Kirsch. *Integration of Programming and Learning in a Control Language for Autonomous Robots Performing Everyday Activities.* PhD thesis, Technische Universität München, 2008.

[Klein 04] G. Klein, J. M. Woods D. D.and Bradshaw and P. J. Hoffman R. R.and Feltovich. *Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity.* IEEE Intelligent Systems, vol. 19, no. 6, pages 91–95, Nov-Dec 2004.

[Kulić 05] Dana Kulić and Elizabeth A. Croft. *Safe planning for human-robot interaction.* Journal of Robotic Systems, vol. 22, no. 7, pages 383–396, 2005.

[Kulic 07] D. Kulic and E. Croft. *Pre-collision safety strategies for human-robot interaction.* Autonomous Robots, vol. 22, no. 2, pages 149–164, 2007.

[Lee 01] P. U. Lee and B. Tversky. *Costs of Switching Perspectives in Route and Survey Description.* In Proceedings of the twenty-third Annual Conference of the Cognitive Science Society, Edinburgh , Scotland, 2001.

[Lochbaum 98] K. E. Lochbaum. *A Collaborative Planning Model of Intentional Structure.* Computational Linguistics, vol. 24, no. 4, pages 525–572, 1998.

[Madhava Krishna 06] K. Madhava Krishna, R. Alami and Simeon T. *Safe proactive plans and their execution.* Robotics and Autonomous Systems, vol. 54, no. 3, pages 244–255, March 2006.

[Marin 08] Luis Marin, Emrah Akin Sisbot and Rachid Alami. *Geometric Tools for Perspective Taking for Human-Robot Interaction.* In Mexican International Conference on Artificial Intelligence (MICAI 2008), Mexico City, Mexico, October 2008.

[Montreuil 07] V. Montreuil, A. Clodic and R. Alami. *Planning Human Centered Robot Activities.* In IEEE Int. Conf. on Systems, Man and Cybernetics (SMC), 2007.

[Nokata 02] M. Nokata, K. Ikuta and H. Ishii. *Safety-optimizing method of human-care robot design and control.* In IEEE International Conference on Robotics and Automation, ICRA, volume 2, pages 1991–1996, Washington DC, USA, May 2002.

[Nonaka 04] S. Nonaka, K. Inoue, T. Arai and Y. Mae. *Evaluation of human sense of security for coexisting robots using virtual reality. 1st report: evaluation of pick and place motion of humanoid robots.* In IEEE International Conference on Robotics and Automation, ICRA, volume 3, pages 2770–2775, New Orleans, USA, April 2004.

[Pollack 03] Martha E. Pollack, Laura Brown, Dirk Colbry, Colleen E. McCarthy, Cheryl Orosz, Bart Peintner, Sailesh Ramakrishnan and Ioannis Tsamardinos. *Autominder: An Intelligent Cognitive Orthotic System for People with Memory Impairment.* Robotics and Autonomous Systems, no. 44, pages 273–282, 2003.

[Rich 97] C. Rich and C. L. Sidner. *COLLAGEN: When Agents Collaborate with People.* Proceedings of the first international conference on Autonomous Agents, 1997.

[Sidner 05] C. L. Sidner, C. Lee, C. Kidd, N. Lesh and C. Rich. *Explorations in Engagement for Humans and Robots.* Artificial Intelligence, vol. 166, no. 1-2, pages 140–164, 2005.

[Siméon 01] T. Siméon, J.-P. Laumond and F. Lamiraux. *Move3D: A generic platform for path planning.* In IEEE International Symposium on Assembly and Task Planning, ISATP, pages 25–30, Fukuoka, Japan, May 2001.

[Sisbot 07a] E. Akin Sisbot, Luis F. Marin Urias, Rachid Alami and Thierry Siméon. *Spatial Reasoning for Human-Robot Interaction.* In IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, San Diego, CA, USA, November 2007.

[Sisbot 07b] Emrah Akin Sisbot, Luis F. Marin-Urias, Rachid Alami and Thierry Simeon. *Human Aware Mobile Robot Motion Planner.* IEEE Transactions on Robotics, vol. 23, pages 874–883, 2007.

[Tambe 97] M. Tambe. *Towards Flexible Teamwork.* Journal of Artificial Intelligence Research, vol. 7, pages 83–124, 1997.

[Tenorth 08] Moritz Tenorth and Michael Beetz. *Towards Practical and Grounded Knowledge Representation Systems for Autonomous Household Robots.* In Proceedings of the 1st International Workshop on Cognition for Technical Systems, München, Germany, 6-8 October, 2008.

[Tomasello 07] Michael Tomasello and Malinda Carpenter. *Shared intentionality.* Developmental Science, vol. 10, no. 1, pages 121–125, 2007.

[Traver 00] V. J. Traver, A. P. del Pobil and M. Perez-Francisco. *Making service robots human-safe.* In IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, volume 1, pages 696–701, Takamatsu, Japan, November 2000.

[Tversky 99] B. Tversky, P.Lee and S. Main Waring. *Why do Speakers Mix perspectives.* Spatial Cognitive Computation, vol. 1, pages 312–399, 1999.

Annex 1: An experimental protocol for Human Robot cooperative manipulation

As a result of preliminary work between LAAS and MPG and a visit of two persons from LAAS to Leipzig, a protocol for studying and analysing human-robot interaction in CHRIS context, has been discussed in the CHRIS internal workshop on cooperation held October 2008.

The scenario is based on a set up similar (see §3.2) to one mentioned in D3 (Systems Engineering Specification).

Depending on the situation and on the available ability when the test will be performed, we will use autonomous robot capabilities or a Wizard-of-Oz software environment that allows to control the robot at or action or task-level (compared to classical joystick, or scripted operation).

The scenario involves H (a Human) & R (a robot). Several aspects can be dealt with:

1. Study of task performance when an individual is:
   - Performing a task alone
   - Performing a task in presence
   - Performing a task in Interaction with another individual
2. Knowledge of the task. We assume that the basic manipulation abilities (actions) are available. One can run examples where:
   - the task is known
   - the task is only partially unknown (from R point of view)

   Task Plan (i.e. a way to achieve the task in a given context). The case is of interest if several possible plans are available. The criterion will be linked to the choices made when elaborating or selecting a plan, and also on conditions that can cause a change of plan
3. Situations will be also examined where:
   - H can do it alone .. but R proposes to help
   - Task cannot be done by H alone
4. Roles of H & R will be also examined.
   - Team-mates (peers)
   - Boss-assistant (with limited initiative)
   - Expert / Novice
5. The agreement process
   - On the task
   - On the roles allocation
   - On the plan (recipe)

Annex 2: Applicable documents

- CHRIS (Cooperative Human Robot Interaction Systems DOW), "Annex I - Description of Work", October 2007.
- CHRIS Deliverable D3 "Systems Engineering Specification", October 2008.
- YARP, description and documents http://eris.liralab.it/yarp/
- LAAS Openrobots software tools: http://softs.laas.fr/openrobots/wiki/

Appendix 3:

Working Paper: Perspective taking in human-robot interactions

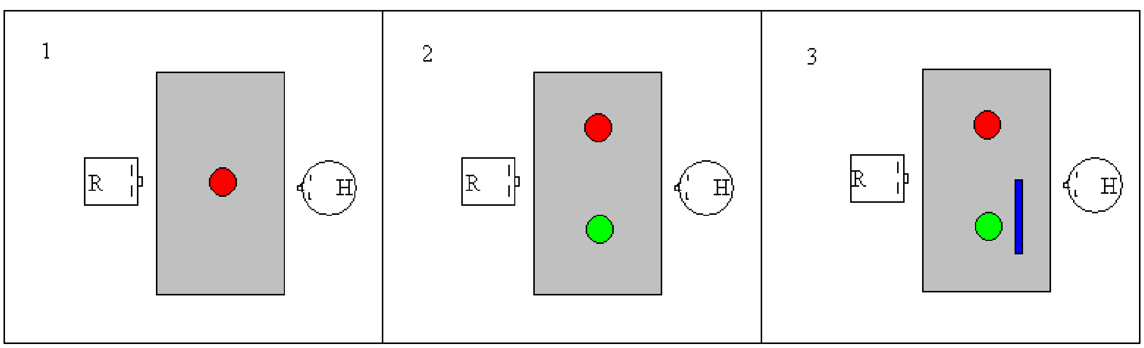
Katharina Hamann, Jasmin Steinwender, Felix Warneken
Max-Planck-Institute for Evolutionary Anthropology, Leipzig, Germany

June 8, 2009

Working Paper: Perspective taking in human-robot interactions

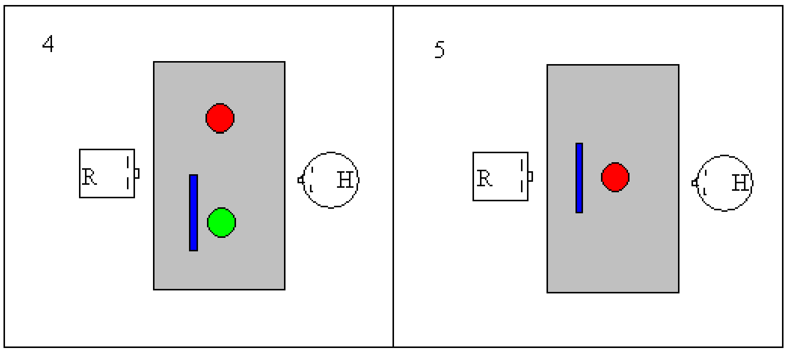
Situations with one person making a request that is ambivalent, i.e., that has two or more possible referents, can be disambiguated in several ways. The most important ones in this context are visual perspective taking and further communication (like counter questions).
One rule considering the speaker's need for efficiency and comfort is that in general, problems that **can** be solved without much communication also **should** be solved without much communication. There might be a range from problems easily solvable that are encountered just nonverbally (on the fly) to problems only possible to be disambiguated verbally.

More specifically, perspective taking abilities might make verbal enquiries unnecessary. To illustrate this point, think of the following scenarios (see also Moll & Tomasello, 2006, and Trafton, Schultz, Bugajska, & Mintz, 2005):



Consider the human (H) asks the robot (R) to hand him the ball. **Situation (1)** has no ambiguity whatsoever and needs no further specifications or abilities. In **situation (2),** however, there are two balls, and as long as there are no other cues (e.g., H points to the ball he wants to have or looks at it pervasively), the only thing the robot can do is to ask ("Do you want the green one or the red one?"). In **situation (3)** the same two balls are on the table, both visible to R just as in (2). *Without* level 1 perspective taking skills, the situation would remain ambiguous, and R again would have to ask. If, however, these skills exist, R can infer that H refers to the red ball from knowing that the barrier prevents H from seeing the green ball. Hence, it should hand the red one over (cf. Trafton et al., 2005). Conversely, if H does not ask R to give him the ball but instead *looks for* a ball and asks for help, R should be able to infer that the ball H can see is probably not the one he is looking for, and thus should give him the green one (cf. Moll & Tomasello, 2006). Note that the two different questions evoke two different actions, but the same level 1 perspective taking skill is necessary to solve either.

To more or less complete the row of possible scenarios with one or two balls on a table, see the following figures:



If the robot is asked by the human to hand over the ball in **Situation (4)**, there is no ambiguity for it as far as it does not "know" about the green ball behind the barrier. Thus, this scenario is similar to (1). **Situation (5)** (see Trafton et al., 2005), instead, poses a problem that R can solve by applying level 1 perspective skills: When asked, it can infer that it probably cannot see something that the human can see. Removing the barrier or moving around it (that is, actions that provide R with H's perspective) would be adequate reactions.

The following table provides an overview of the things just described. First 2 rows: Depending on the existence of level 1 perspective taking skills, the situations either need to be disambiguated verbally (=1) or not (=0). Moreover, in parentheses the **success rates** expected (ideally) are outlined – in case that verbal disambiguation is *not* available (for with it, the expected success would be 100% in every situation). The third and fourth rows display **reaction times (RT)** to be expected with or without level 1 perspective taking skills, and with the possibility of communication at hand. Note that communication is not considered to be used if perspective taking is possible.

	Situation 1	Situation 2	Situation 3	Situation 4	Situation 5
Level 1 perspective taking	0 (100%)	1 (50%)	**0 (100%)**	0 (100%)	**0 (100%)**
No perspective taking ability	0 (100%)	1 (50%)	**1 (50%)**	0 (100%)	**1 (0%)**
Reaction time expected *with* L1	Fast	Slow	**Fast**	Fast	**Fast**
Reaction time expected *without* L1	Fast	Slow	**Slow**	Fast	**Very slow**

As mentioned, verbal disambiguation would always lead to success in the end. But back-and forth questioning takes its time, and thus success occurs at the expense of time (and human nerves). Situation (5) might be particularly time consuming in that more communication might be necessary than in scenarios (2) and (3) (the human would have to tell the robot to remove the barrier first or move around the table, and then perhaps repeat his request). Perspective taking abilities, in contrast, allow for more efficient and user-friendly actions of the robot.

Example of Level 2 perspective taking

Perhaps it is useful to imagine that the robot for some reason “knows” about the ball behind the barriers in (4) and (5). This might be of importance if his visual data input comes from cameras providing it with a complete map of the room (for instance, if they are attached to the ceiling). Now situation (5) does not pose any problem anymore. But (4) becomes tricky, especially if the human doesn’t know about the robot overlooking the whole situation. To react adequately, R now has to take into account that H does not expect R to see the green ball and thus refers to the red one. Thus, it must consider that H sees the green ball as being invisible for R – that is, in some sense, *how* he sees it. This poses a more complex level 2 perspective problem.

Other tasks than visual perspective taking

So far, the scenarios contained visible/invisible objects. One could also imagine other tasks, like, e.g., proximity/distance problems. The robot would have to understand that objects can be seen much better from proximity than from distance. It would be interesting to know if the very same algorithms would hold for several of these perspective taking tasks (with only changing the key variables).

Suggestion for a decision tree

When faced with a situation of the kind just described, the robot might follow a decision path like that:

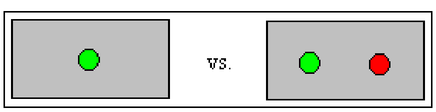
1. Can I solve it on the fly (no ambiguity of question & situation)? – If yes, solve it. If no,
2. Is there a perspective problem? – If no, ask (disambiguate verbally). If yes,
3. Try to disambiguate by applying perspective taking skills. – If possible, solve it. If impossible,
4. Ask (disambiguate verbally).

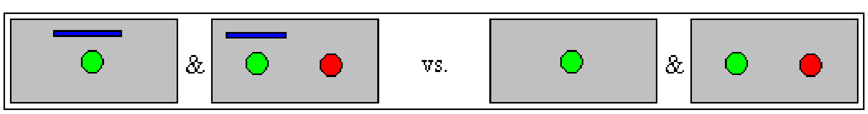
Further considerations

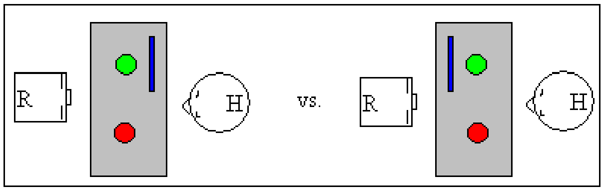
The decision tree just outlined requires a robotic system disposing of (at least) the following skills: verbal skills, level 1 perspective taking skills, and abilities of visual and auditory perception (that are just taken for granted here). To test the expectations about the effectiveness and efficiency of perspective taking abilities one could compare the outcomes of systems with and without these skills. The expectations about reaction times and user friendliness could be incorporated.

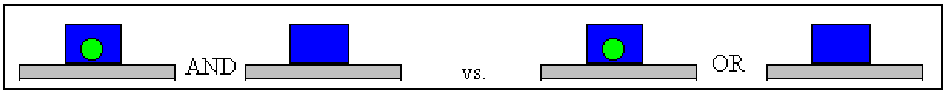
The different situations described above have additional components that can be varied systematically when experiments are to be conducted. A list of all variables (and their values) follows:

1. Number of objects (1 vs. 2)
2. General visibility of objects (barrier vs. no barrier)
3. Side of object (in the front vs. behind a barrier from R’s perspective)
4. Perspective taking ability (yes vs. no)
5. Communication (yes vs. no)

1.


2.


3. 

4. 

5. [No diagram]

References:

Moll, H. & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology*, *24*, 603-613(11).

Trafton, J., Schultz, A., Bugajska, M., & Mintz, F. (2005). Perspective-taking with robots: experiments and models. In *IEEE International Workshop on Robot and Human Interactive Communication, 2005. ROMAN 2005*, (pp.580-584).