Marie Curie Individual Fellowship
FP7 Program

Report 3:

Scenarios and Validation Tasks

Produced by:	Raquel Ros Espinoza
Grant agreement Number:	220368
Project title:	Adapting Robot Behavior based on Interaction
Project acronym:	ARBI
Project starting date:	01 August 2008
Project ending date:	31 July 2008
Project coordinator:	Dr. Rachid Alami

Toulouse, August 2010.

1 Architecture Overview

In the current version the whole system is integrated into two robotic platforms as components of the LAAS architecture [1]: HRP-2 and Jido. In order to acquire and keep a coherent model of the environment, a number of additional modules are used:

- *Object Recognition Module*: This visual recognition and localization module detects objects around the robot by detecting special markers. Once a marker is detected, the identification of the object attached to this marker as well as its position are provided.
- *Human Detection Module*: As the human looking direction is crucial for the system, this data is collected from the high precision motion capture cameras. A specially marked helmet placed on human's head is tracked and the 3D position as well as the orientation are provided.
- *Robot Manager Module*: This module serves as a provider of robot's current configuration.
- *Geometric Reasoning*: As explained in Report 2, this module is in charge of computing the geometric reasoning capabilities of the robot. The inputs to this module are computed by the above three modules. Based on this information, the 3D model of the world is obtained.

Additional modules outside the LAAS architecture, but linked to it, are:

- *Dialog*: Processes natural language in order to represent it using the model of the robot's knowledge.
- *ORO*: Robot's symbolic knowledge.
- *Clarification Module*: This module is directly linked with the dialog module. It is in charge of grounding the referents (disambiguation), as well as generating non ambiguous descriptions.
- *Decisional Reasoning*: In charge of the decisional framework of the robot and the enhancement of the human-robot interaction.

We must remark that during the project the different components have been developed in different periods. Therefore, the architectures used in the scenarios we report on next may vary from the one just described.

2 Clarification Scenarios - *Which one?*

In order to test our approach we have defined a set of game scenarios where ambiguities may easily arise during an interaction. In all cases there is a face-to-face interaction between a human and the robot, both in simulation and in real world. Both agents are around a table with different objects on it. The human

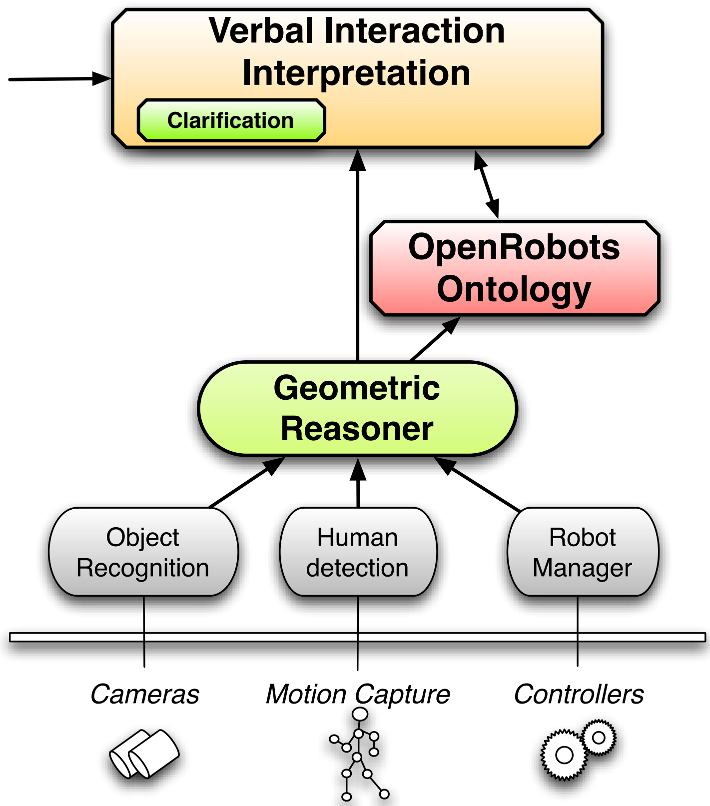


Figure 1: The general architecture of the system. Perspective reasoner constantly updates environment information and answers to the queries of clarification module.

asks the robot to hand her an object and the robot has to infer which object the human is referring to, either by itself or asking for additional information. Figure 2 shows the graphic interface of the Move3D software platform for each scenario. The available objects in the different scenarios are defined based on three attributes:

- shape: cube, box and sphere.
- color: red, green and blue.
- size: small and large.

This information is so far initially given by hand to the system. There is an additional object in the scenario, a teddy bear. It is used to occlude objects from the human perspective to test the visual perspective taking skill of our system.

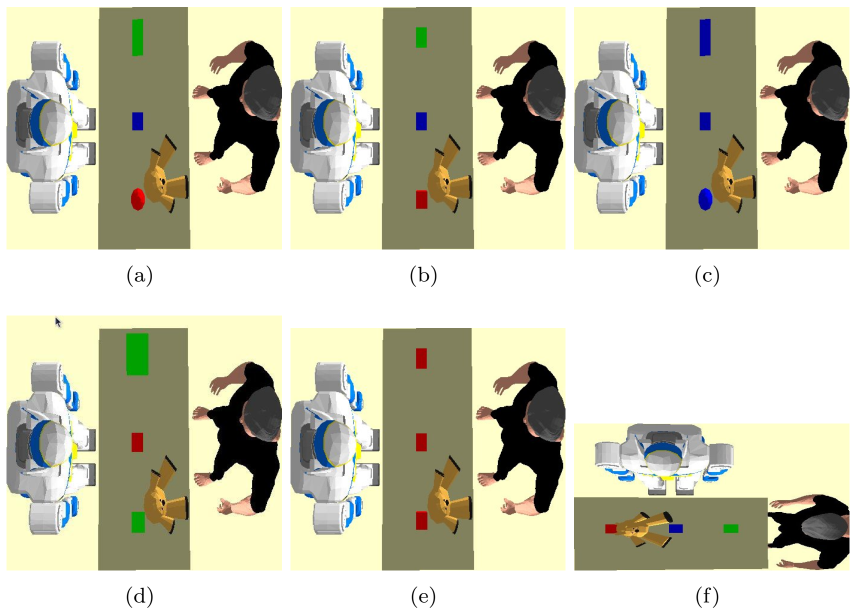


Figure 2: Simulation scenarios.

The layout of the scenarios is not random, and in fact it tries to cover a wide variety of situations where clarification should be performed in different ways. Table 1 describes each scenario in terms of the nature of the ambiguity (if present) and the possible discriminative descriptors that could be used to solve them (if possible).

The first scenario (Figure 2a) corresponds to the "less" ambiguous scenario, since only one attribute would cause confusion, i.e. size. Hence, unless the human asks for "the small object", the statement itself would be clear an no specific clarification would be needed. The second and third scenario (Figure 2b and Figure 2c) are a bit more tricky, since in each case, there are two ambiguous features, shape/size and color/size, respectively. Given the query "Give me

Scn	Ambiguity	Unambiguous descriptors
1	size	vpt $\wedge$ (color $\vee$ shape $\vee$ loc wrt H or R)
2	shape, size	vpt $\wedge$ (color $\vee$ loc wrt H or R)
3	color, size	vpt $\wedge$ (shape $\vee$ loc wrt H or R)
4	shape, color, size	vpt $\wedge$ (size $\vee$ loc wrt H or R)
5	shape, color, size	vpt $\wedge$ loc wrt H or R
6	shape, size, loc wrt H	vpt $\wedge$ (color $\vee$ loc wrt R)

Table 1: Scenarios' description based on ambiguities sources and their correspondent solutions (vpt stands for Visual Perspective Taking).

the cube" or "Give me the blue object" in each scenario, applying only visual perspective taking would only eliminate one candidate object (the occluded object from the human perspective). Therefore, one additional query to the human is required in order to find out the correct object. Any of the remaining descriptors could be applied.

Scenario 4 (Figure 2d) is an interesting example where, based on the human's query, there might not be any confusion at all or an ambiguity takes place. In the latter case, the ambiguity may be solved either immediately or with an additional clarification. Suppose that the human asks for "the large box". In this situation there is no ambiguity at all since there is only one large object on the table. However, using the same attribute, i.e. size, but with a different value, does generate an ambiguity. This way, when the human asks either for "the small object" or for "the green box" (switching the feature to color), using visual perspective taking the robot can eliminate the occluded object (the small green box) clarifying the ambiguity by itself. Finally, additional information should be asked in the case where the query is with respect to "the box". In this case, after applying visual perspective taking, there are still two possible objects: the large green box and the red small one. Hence, the system should clarify with the human using any of the available unambiguous descriptors, i.e. color, size or location with respect to any of the agents.

The two remaining scenarios describe situations where spatial perspective has a role, both, in solving an ambiguity, and in creating one. Scenario 5 (Figure 2e) corresponds to the former case. The objects on the table are all the same, and only their location varies. Asking for "the red small box" does not provide enough information since even using visual perspective taking, there are still two available solutions. Thus, the human should provide a spatial location when referring to the object. If the referent object is missing, then the robot should ask for it in order to find out which object the human is referring to. Finally, the last scenario (Figure 2f) describes a situation where the objects are in the same location from the human point of view, i.e. in front. Thus, the system cannot use this descriptor to clarify the query "give me the box in front of me". Instead, it should ask for the color, i.e. blue or green, since the red box is not visible for the human, or the location with respect to its position, i.e. its front or its left.

We have adapted two of the simulation scenarios to the real environment. More precisely, we use bottles as objects to reference and a large box as an obstacle to occlude objects. The first test configuration, Figure 3, corresponds the the second simulation scenario, where all objects have the same shape (bottles) but different colors. The second, Figure 4, corresponds to the last simulation scenario, where the spatial description of the objects is ambiguous from the human point of view, but not from the robot's. Instead of modifying the human's position, we have moved the objects on the table. This way we could also verify the correct computation of the spatial positions of the objects with respect to both, the robot and the human.

Figures 3d and 4d illustrate the robot's representation of the world state shown in the pictures above. Figures 3e and 4e depict the focus of attention of

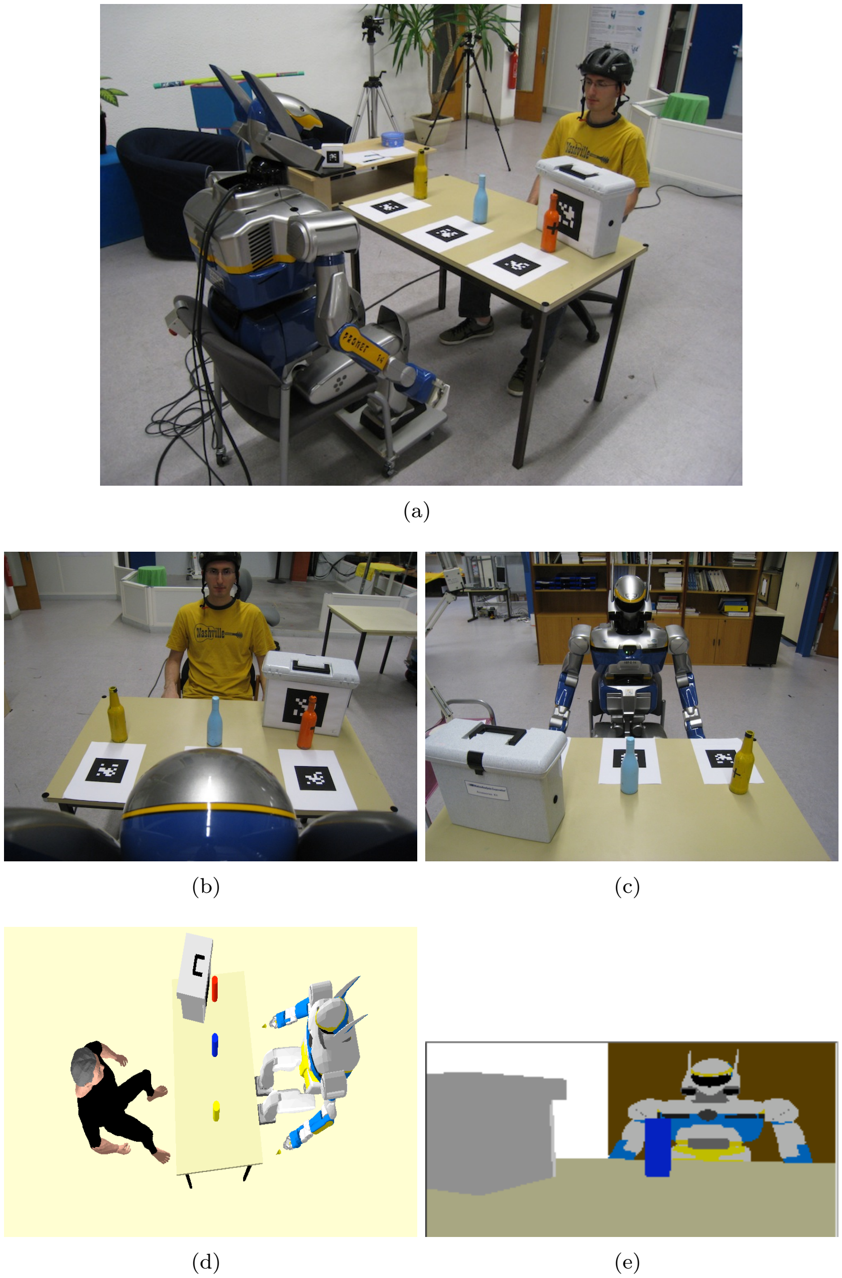
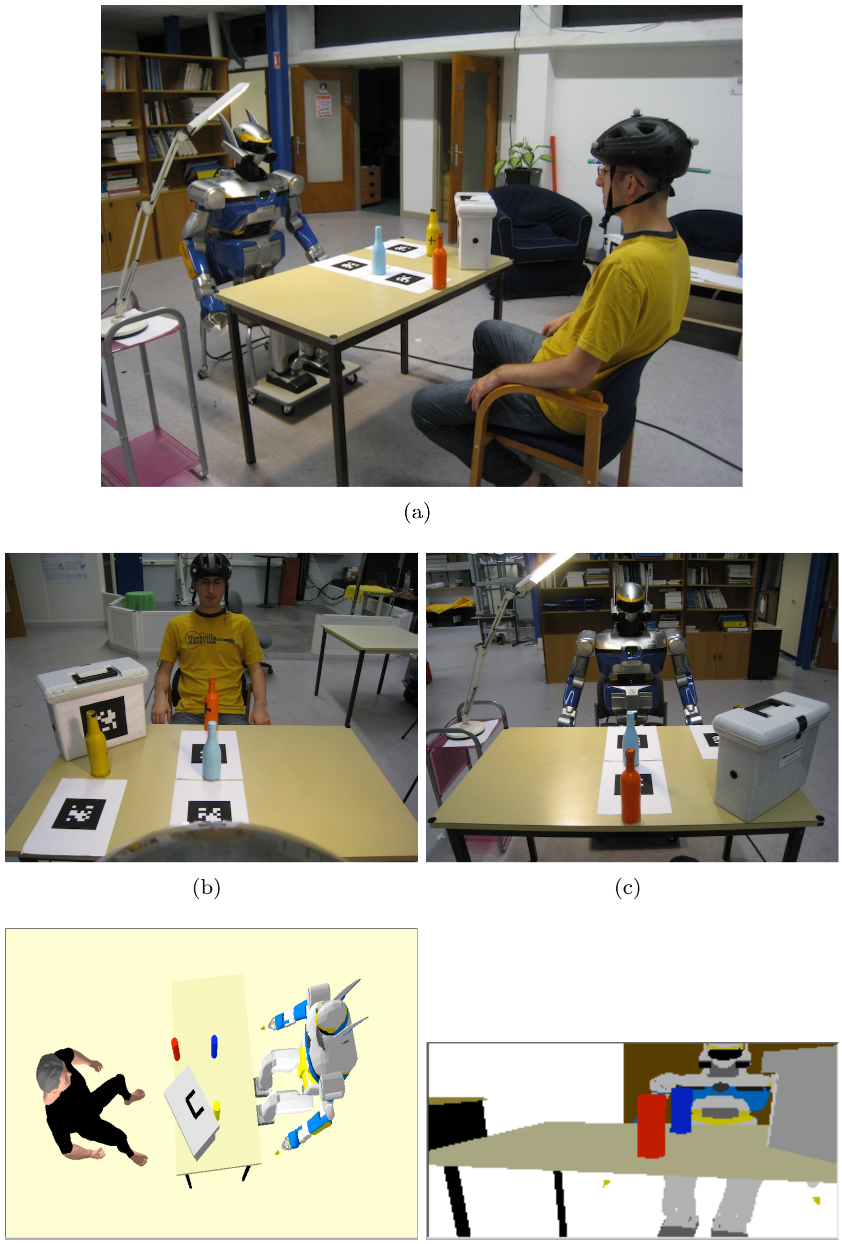


Figure 3: Bottles scenario 1: (a) real scenario global view, (b) human's perspective, (c) robot's perspective, (d) Move3D global scenario (e) Move3D human's perspective.



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

Figure 4: Bottles scenario 2: (a) real scenario global view, (b) human's perspective, (c) robot's perspective, (d) Move3D global scenario (e) Move3D human's perspective.

Id	Type	Color	Shape	Location
purple-ball	**ball**	**purple**	sphere	on $shelf_1$
orange-bottle	**Bottle**, Tableware	**orange**	cylinder	on *big_table*
blue-bottle	**Bottle**	**blue**	cylinder	on *big_table*
red-bottle	**Bottle**	**red**	cylinder	in *trashbin*
orange-box	**GameObj**, Object	**orange**	**cube**	on *accesskit*

Id	Spatial PT		Visual PT	
	Robot	Human	Robot	Human
purple-ball	back near	front far	false	false
orange-bottle	front right near	front left near	true	false
blue-bottle	front left near	front near	true	true
red-bottle	front right far	back left near	false	false
orange-box	front near	front left near	true	true

Table 2: Example of objects' descriptions. In bold, information given at initialization. All other descriptors are automatically computed by the geometric reasoning or inferred from the ontology on-line.

the human based on the robot's perception. Note that the images only show the current focus of attention of the human, but not the complete human's field of view. Thus, Figure 3d corresponds to a snapshot where the human was looking towards his left, observing the obstacle and the blue bottle only. From this view, we can clearly observe that the yellow bottle is not visible for him, even if its focus of attention is pointing at such direction. In Figure 3e, we can see that the human is able to see the two aligned bottles in front of him and part of the obstacle.

The outcome of this first work has been published in [5], where ORO was not yet implemented (we used a plain data base instead). The next step was then to include ORO and interact with the ontology. An example of the information stored in the ORO is shown in Table 2. We indicate wether the description is manually given or computed. For example, the information provided for the object *orange-bottle* is that it belongs to the class Bottle and that its color is orange. Based on the knowledge in the common sense ontology, the robot can infer it is also a Tableware, and that its shape is cylindrical. The remaining descriptors (location, spatial and visual perspective taking) are automatically computed through geometric reasoning. Figure 5 describes three situations where the human participation of the human and the type of information provided is varied. This work has been published in [4].

3 Tell me what you know about...? and Naming

We have designed an interactive task that exploits the robot's knowledge while using the different mechanisms presented in this work. The scenario for the task

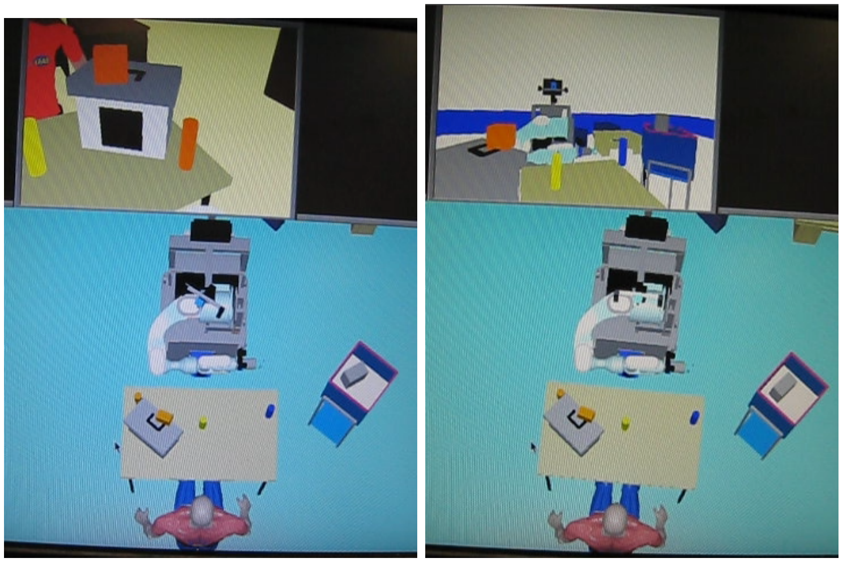
human	Give me the ball.
robot	*(Obtaining possible objects: objectL = [purple-ball, orange-ball])* *(Removing non visible objects for the human: objectL = [orange-ball])* You want the orange-ball (and eventually it would handle it to the human).
human	Give me the bottle.
robot	*(Obtaining possible objects: objectL = [blue-bottle, yellow-bottle, orange-bottle, red-bottle])* *(Removing non visible objects for the human: objectL = [blue-bottle, yellow-bottle])* *(Obtaining complete discriminants: color.)* Which one: blue or yellow?
human	Blue.
robot	*(Obtaining possible objects: objectL = [blue-bottle])* You want the blue-bottle (and eventually it would handle it to the human).
human	Give me the box.
robot	*(Obtaining possible objects: objectL = [orange-box, accesskit, white-box])* *(Removing non visible objects for the human: objectL = [orange-box, accesskit, white-box])* *(Obtaining complete discriminants: none (all boxes share type, shape and partially color –one orange and two white–))* Which one? I need more information.
human	On your left.
robot	*(Obtaining possible objects: objectL = [white-box])* You want the white-box (and eventually it would handle it to the human).

Figure 5: “Which one?” example. The uppermost cell includes an ambiguous situation solvable through visual perspective taking, the middle cell a feature-based solution, the lower cell a spatial perspective taking based solution.

consists in a face-to-face interaction around a table with objects. The human may ask the robot the following questions:

Where is the ***object_description*****?**

The robot indicates the location of the object based on spatial perspective taking and symbolic location descriptors.



(a) (b)

Figure 6: Computing visibility of the orange bottle for (a) the robot and (b) the human. On top we show the visual perspective for the agent under evaluation, while on the bottom, an overview of the environment

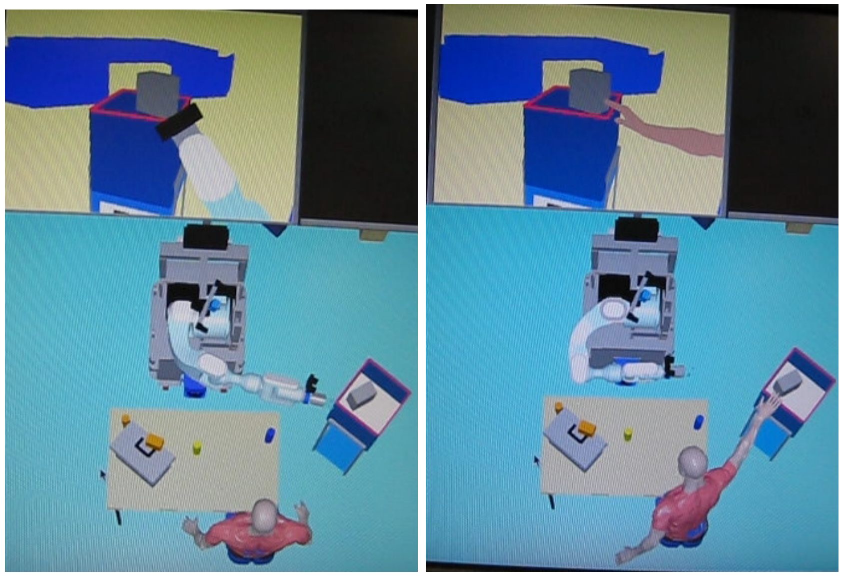
Is the *object_description* visible?

The robot computes the visibility of the object from both agents' perspectives (robot and human) and indicates whether the object is visible or not. If it is, it also indicates if it is directly visible (within the agent's current FOV) or if the object is visible by turning the head (out of FOV). The view of the agents is displayed in the screen at the same time (Figures 6).

Is the *object_description* reachable?

The robot computes the object's reachability from both agents' perspectives (robot and human) and indicates whether the gray box is reachable or not. The screen displays the movement of the agent when computing the reaching posture (Figure 7).

From the human query, the decisional reasoner extracts the description of the referred object. The description can be either the id (eg. YELLOW_BOTTLE) or a set of attributes about the object (eg. yellow bottle). In the first case, if the id corresponds to a known object, the robot can directly answer the question. On the contrary, if the object is unknown, then a learning phase takes place, where the human describes recursively the type of the object until a known type is reached. Figure 8 shows an example of the learning process where the robot asks for the type of object until recognizing a known type. In the second



(a) (b)

Figure 7: Computing reachability for (a) the robot and (b) the human. On top we show a narrow view of the movement to reach the object, while on the bottom, an overview of the environment

case, where attributes describing the object are given, the reasoning engine has to figure out the id of the object. However, the information provided by the human may be incomplete or ambiguous. Therefore, the robot must ask the human additional information until identifying the object using the clarification algorithm. Figure 9 illustrates different interaction scripts where the robot finds out the referred object and answers the human query. This work has been published in [6].

4 Spy Game

This game is based on the traditional children game "I Spy". The idea is to discover the object (which should be visible) or concept one of the participants is thinking of by asking questions such as: "Is it green? Is it a machine? Is it on your left?", etc. When playing, children exploit their knowledge about the world while categorizing and describing objects through useful discriminants.

The Spy Game (see Figure 10 for an example) starts with the human user giving a first hint that corresponds to the initial description indicated in the clarification algorithm introduced in Report 2. The robot tries to figure out the object automatically searching for discriminants (complete or partial) and querying the human for their values until either discovering the object or giving

human	(Shows an object to the robot).
robot	(Identifies an unknown object.) What's the name of the object?
human	mist
robot	What kind of object is mist?
human	laptop
robot	What is a laptop?
human	computer
robot	What is a computer?
human	object
robot	Now I know what mist is.

Figure 8: Learning process example where the robot learns that mist is a computer.

human	Where is the bottle?
robot	*(Available object ids: BLUE_BOTTLE, YELLOW_BOTTLE and ORANGE_BOTTLE.)* *(Complete discriminants:* hasColor) Which color is the object?
human	Yellow.
robot	The YELLOW_BOTTLE is located: - from my perspective: FRONT - from your perspective: LEFT - on HRP2TABLE.
human	Is the orange bottle visible?
robot	*(Available object ids: ORANGE_BOTTLE)* The ORANGE_BOTTLE is visible for me (out of FOV) and not visible for you.
human	Is the white box on my left reachable?
robot	*(Available object ids: SPACENAVBOX)* The SPACENAVBOX is reachable for me and reachable for you.

Figure 9: Interactive task example.

up (if no object was found with the final description). An example of the dialog while guessing the object is shown in Figure **??**This work has been published in [4, 3].

5 Other

It is worth to remark that some of the work produced in this has taken part of the CHRIS project, in where I have actively taken part during this time. Moreover,

Human	It is a tableware.
Robot	*(Obtaining possible objects: objectL = [blue-bottle, yellow-bottle, orange-bottle, red-bottle and cup-with-handle])*
	(Removing non visible objects for the human and the robot: objectL = [blue-bottle, yellow-bottle, cup-with-handle])
	(Obtaining discriminants: type, color.)
	Which type of object is: bottle or cup?
Human	Bottle.
Robot	*(Obtaining possible objects: objectL = [blue-bottle, yellow-bottle])*
	[Obtaining discriminants: color.]
	What color the object is: blue or yellow?
Human	Blue.
Robot	*[Obtaining possible objects: objectL = [blue-bottle])*
	The object is the blue bottle!

Figure 10: Spy Game example.

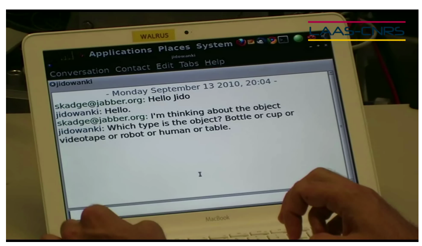


Figure 11: Snapshot while chatting with the robot in the Spy Game scenario.

after an integration activity in year two of the project, a joint publication has been carried out, where all the partners of the consortium have participated. In this work, [2], the naming scenario has been successfully integrated into the BERT2 platform at the Bristol Robotics Laboratory. Figure 12 shows a snapshot of the naming validation task.

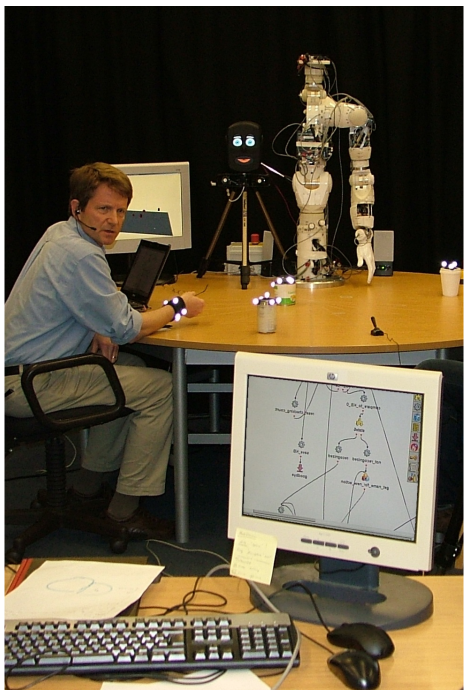


Figure 12: Integration at Bristol Robotics Laboratory

6 Publications

The outcome of the work realized during this period has produced the following publications and participations:

- E. A. Sisbot, R. Ros and R. Alami. *Situation Assessment for Human-Robot Interaction.* Submitted to ICRA 2011.
- S. Lemaignan, R. Ros, L. Mösenlechner, R. Alami, and M. Beetz. *ORO, a knowledge management module for cognitive architectures in robotics.* In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2010. To appear.
- R. Ros, S. Lemaignan, E. A. Sisbot, R. Alami, J. Steinwender, K. Hamann, and F. Warneken. *Which one? Grounding the referent based on efficient human-robot interaction.* In 19th IEEE International Symposium in Robot and Human Interactive Communication (RO-MAN), 2010. To appear. **CoTeSys Award**.
- R. Ros, E. A. Sisbot, S. Lemaignan, A. Pandey, and R. Alami. *Robot, tell me what you know about...?: Expressing robots knowledge through interaction.* In Proceedings of the ICRA 2010 Workshop on Interactive Communication for Autonomous Intel ligent Robots (ICAIR), pages 2629, 2010.
- R. Ros, E. A. Sisbot, R. Alami, J. Steinwender, K. Hamann, and F. Warneken. *Solving ambiguities with perspective taking.* In Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pages 181182. ACM, 2010.
- J. Steinwender, K. Hamann, R. Ros, and E. A. Sisbot. *Human Perspective taking as a model for solving ambiguities in HRI.* Oral presentation in Developmental Psychology Contributions to Human Robot Cooperation Workshop at Humanoids 2009.

We are currently working on a journal paper in order to describe the overall approach at the current stage.

References

[1] R. Alami, R. Chatila, S. Fleury, M. Ghallab, and F. Ingrand. An architecture for autonomy. *International Journal of Robotics Research*, 17:315–337, 1998.

[2] S. Lallée, S. Lemaignan, A. Lenz, C. Melhuish, L. Natale, S. Skachek, T. van Der Zant, F. Warneken, and P. F. Dominey. Towards a platform-independent cooperative human-robot interaction system: I. perception. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010. To appear.

[3] S. Lemaignan, R. Ros, L. Mösenlechner, R. Alami, and M. Beetz. Oro, a knowledge management module for cognitive architectures in robotics. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010. To appear.

[4] R. Ros, S. Lemaignan, E. A. Sisbot, R. Alami, J. Steinwender, K. Hamann, and F. Warneken. Which one? grounding the referent based on efficient human-robot interaction. In *19th IEEE International Symposium in Robot and Human Interactive Communication*, 2010. To appear.

[5] R. Ros, E. A. Sisbot, R. Alami, J. Steinwender, K. Hamann, and F. Warneken. Solving ambiguities with perspective taking. In *Proceedings of the 5th ACM/IEEE international conference on Human-Robot Interaction*, pages 181–182. ACM, 2010.

[6] R. Ros, E. A. Sisbot, S. Lemaignan, A. Pandey, and R. Alami. Robot, tell me what you know about...?: Expressing robot's knowledge through interaction. In *Proceedings of the ICRA 2010 Workshop on Interactive Communication for Autonomous Intelligent Robots (ICAIR)*, pages 26–29, 2010.