A management of mutual belief for human-robot interaction

Aurélie Clodic†, Maxime Ransan†, Rachid Alami†, Vincent Montreuil†‡

***Abstract*— Human-robot collaborative task achievement requires the robot to reason not only about its current beliefs but also about the ones of its human partner. In this paper, we introduce a framework to manage shared knowledge for a robotic system dedicated to interactive task achievement with a human. In a first part, we define which beliefs should be taken into account ; we then explain a manner to achieve them using communication schemes. Several examples are presented to illustrate the purpose of beliefs management including a real experiment demonstrating a "give object" task between the Jido robotic platform and a human.**

I. INTRODUCTION

We consider that service robots, such as Jido or Rackham (figure 1) need to be interactive and able to answer a number of requests from humans. For instance they can be asked to perform an action, participate in a collaborative task, enumerate their capacities and be pro-active i.e. being able to propose to achieve a task when detecting a relevant context.

It is well established that when people observe and interact with an autonomous robot, they generally apply a social model to it; this has led to the definition of social robots which are the ones requiring such a social model in order to interact and to be understood [1].

One of today key issues in Human Robot Interaction is making the robot and the human understandable and predictable to each other [2]. One well known approach consists in taking human's perspective [3]. Perspective taking can be interpreted at a "geometric" level (adapting robot motion to human presence [4]) and at a "symbolic" level (goals, task achievement process, task and environment state [5], [6]). This paper defines a way to represent human's perspective in a robotic architecture.

Besides human-robot collaborative work in space applications [7] where a remote or high-level control system may be present to organize the work globally (even if after interaction is done without intermediary), we are in a situation where the robot and the human collaborate directly. No proxy will be present between the human and the robot as it is done for example in Machinetta teamwork [8], [9]. In that sense, we are closer to the Collagen approach [10], [11] trying to learn from dialog and language theory.

Consequently, we consider the robot to be an individual agent with its own beliefs, reasoning abilities and perception capacities. This has led us to define knowledge that need to be shared in a particular way.

Moreover, we do not deal with *adjustable autonomy* [12], where the goal is to find the right level of autonomy of the robot and to determine whether and when transfer of control should occur from the agent to other entities. In our case the question is slightly different, since it consists in determining whether and when the robot should interact and/or take initiative towards the human, when they are both trying to achieve a common task. The human and the robot share the same environment, they are often close to each other and perceive each other's activity. The challenge is to equip the robot with suitable context-dependent abilities to make it capable of achieving tasks in the vicinity and/or in interaction with a human partner. We can call such issues *adjustable interaction*.

II. BELIEFS

Joint intention theory (JIT) [13], [14] states that a joint action could not be seen as a collection of individual ones but that agents working together need to share belief. This notion is depicted as mutual belief.

Similar notion could be found in Clark joint action theory [15] as a grounding criterion or common ground: "once we have formulated a message, we must do more than just send it off. We need to assure ourselves that it has been understood as we intended it to be.(...) For whatever we say, our goal is to reach the grounding criterion: that we and our addressees mutually believe that they have understood what we meant well enough for current purposes." However here, a new parameter is inserted which is the understanding of the shared knowledge, i.e. at which point could we be sure an information has not only been perceived but also well understood.

A central notion in collaborative systems is mutual belief (MB) based on the concept of unilateral mutual belief (BMB). From Kumar ([16]), we have:

$$
\begin{aligned}
(BMB\ x\ y\ p) &\triangleq (Bel\ x\ p\ \wedge\ (BMB\ y\ x\ p)) \\
(MB\ x\ y\ p) &\triangleq (BMB\ x\ y\ p)\ \wedge\ (BMP\ y\ x\ p) \\
&\triangleq (Bel\ x\ p\ \wedge\ (BMB\ y\ x\ p))\ \wedge \\
&\quad\ (Bel\ y\ p\ \wedge\ (BMB\ x\ y\ p)) \\
&\triangleq (Bel\ x\ p\ \wedge\ (Bel\ y\ p\ \wedge\ (BMB\ x\ y\ p)))\ \wedge \\
&\quad\ (Bel\ y\ p\ \wedge\ (Bel\ x\ p\ \wedge\ (BMB\ y\ x\ p)))
\end{aligned}
$$

According to this definition, to obtain mutual beliefs of two agents concerning p, we need beliefs of the two agents. This notion could not be used as is in our context because no

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

Authors are from LAAS-CNRS†, Université de Toulouse, 7, avenue du Colonel Roche, 31077 Toulouse Cedex 4, France and Université Paul Sabatier‡, 118, route de Narbonne, 31062 Toulouse, France firstname.name@laas.fr

(a) Jido : a mobile manipulator

(b) Rackham : an interactive museum guide

Fig. 1. Robots Jido and Rackham in interaction context

agent has access to other agent knowledge and belief. Let's consider consider the human h with whom the robot will interact, and r the robot itself. Human beliefs the robot has access to are never $(Bel\ h\ p)$ but $(Bel\ r\ (Bel\ h\ p))$, i.e. access to the human beliefs is done through robot perception. In the same manner, we do not have $(Bel\ h\ (Bel\ r\ p))$ but $(Bel\ r\ (Bel\ h\ (Bel\ r\ p)))$. This could be found in the BMB definition:

$$
\begin{aligned}
(BMB\ x\ y\ p) &\triangleq (Bel\ x\ p\ \wedge\ (BMB\ y\ x\ p)) \\
&\triangleq (Bel\ x\ p)\ \wedge\ (Bel\ x\ (BMB\ y\ x\ p)) \\
&\triangleq (Bel\ x\ p)\ \wedge\ (Bel\ x\ (Bel\ y\ p\ \wedge\ (BMB\ x\ y\ p))) \\
&\triangleq (Bel\ x\ p)\ \wedge\ (Bel\ x\ (Bel\ y\ p))\ \wedge \\
&\quad (Bel\ x\ (Bel\ y\ (Bel\ x\ p\ \wedge\ (BMB\ y\ x\ p)))) \\
&\triangleq (Bel\ x\ p)\ \wedge\ (Bel\ x\ (Bel\ y\ p))\ \wedge \\
&\quad (Bel\ x\ (Bel\ y\ (Bel\ x\ p)))\ \wedge \\
&\quad (Bel\ x\ (Bel\ y\ (Bel\ x\ (Bel\ y\ p\ \wedge (BMB\ x\ y\ p)))))
\end{aligned}
$$

In the rest of this paper we will consider a truncated form that we call UMB for unilateral mutual belief :

$$
\begin{aligned}
(UMB\ x\ y\ p) \triangleq\ & (Bel\ x\ p)\ \wedge\ (Bel\ x\ (Bel\ y\ p))\ \wedge \\
& (Bel\ x\ (Bel\ y\ (Bel\ x\ p)))\ \wedge \\
& (Bel\ x\ (Bel\ y\ (Bel\ x\ (Bel\ y\ p))))
\end{aligned}
$$

This definition as applied in our human-robot interaction context is illustrated in figure 2. It implies that we will give the robot (through perception, dialog abilities and decision process) knowledge concerning:

- $(Bel\ r\ p)$: its beliefs (in one sense, that could be assimilated to knowledge, if we consider that the robot knows its state),
- $(Bel\ r\ (Bel\ h\ p))$: its beliefs concerning human's belief,
- $(Bel\ r\ (Bel\ h\ (Bel\ r\ p)))$: its beliefs concerning human's belief concerning its beliefs,
- $(Bel\ r\ (Bel\ h\ (Bel\ r(Bel\ h\ p))))$, its beliefs concerning human's belief concerning robot beliefs concerning human's beliefs.

Let's consider two examples to illustrate what those beliefs mean in real world robotic applications.

Fig. 3. Rackham exhibiting (or not) its beliefs to person via the use of an external representation

First, suppose Rackham is in a room with the face detection system turned on, meaning it is capable of detecting people's head which are in its camera field of view. A human h, willing to interact with Rackham, comes close to the robot. Let's consider the fact p : $h\ is\ detected\ by\ r$, since the person is detected by the robot, we can conclude the following belief: $(Bel\ r\ p))$. In the case the robot does not give any feedback (and the human does not have any a priori information on the robot abilities) no additional knowledge could be infered from the current situation. If the robot now displays detection information (via the use of an external representation, for example displaying the video stream with a square on the detected face as shown in figure 3), new beliefs can be added, assuming that the information displayed is perceived by the human. The belief $(Bel\ r\ (Bel\ h\ p))$ will be part of the robot knowledge since the human is informed that he has been detected. In that case (because the belief concerns robot's detection) it is equivalent to $(Bel\ r\ (Bel\ h\ (Bel\ r\ p)))$. The last belief is the following: $(Bel\ r\ (Bel\ h\ (Bel\ r\ (Bel\ h\ p))))$, it represents the fact that the human is aware that the robot knows that he knows.

The second example takes place in a guiding context where Rackham and a visitor agree on a destination. In order to reach the desired destination, the robot computes a trajectory p : $trajectoryknown$, implying that we've got $(Bel\ r\ trajectoryknown)$. What we have observed in a previous experience in the Space City Museum ([17], [18])

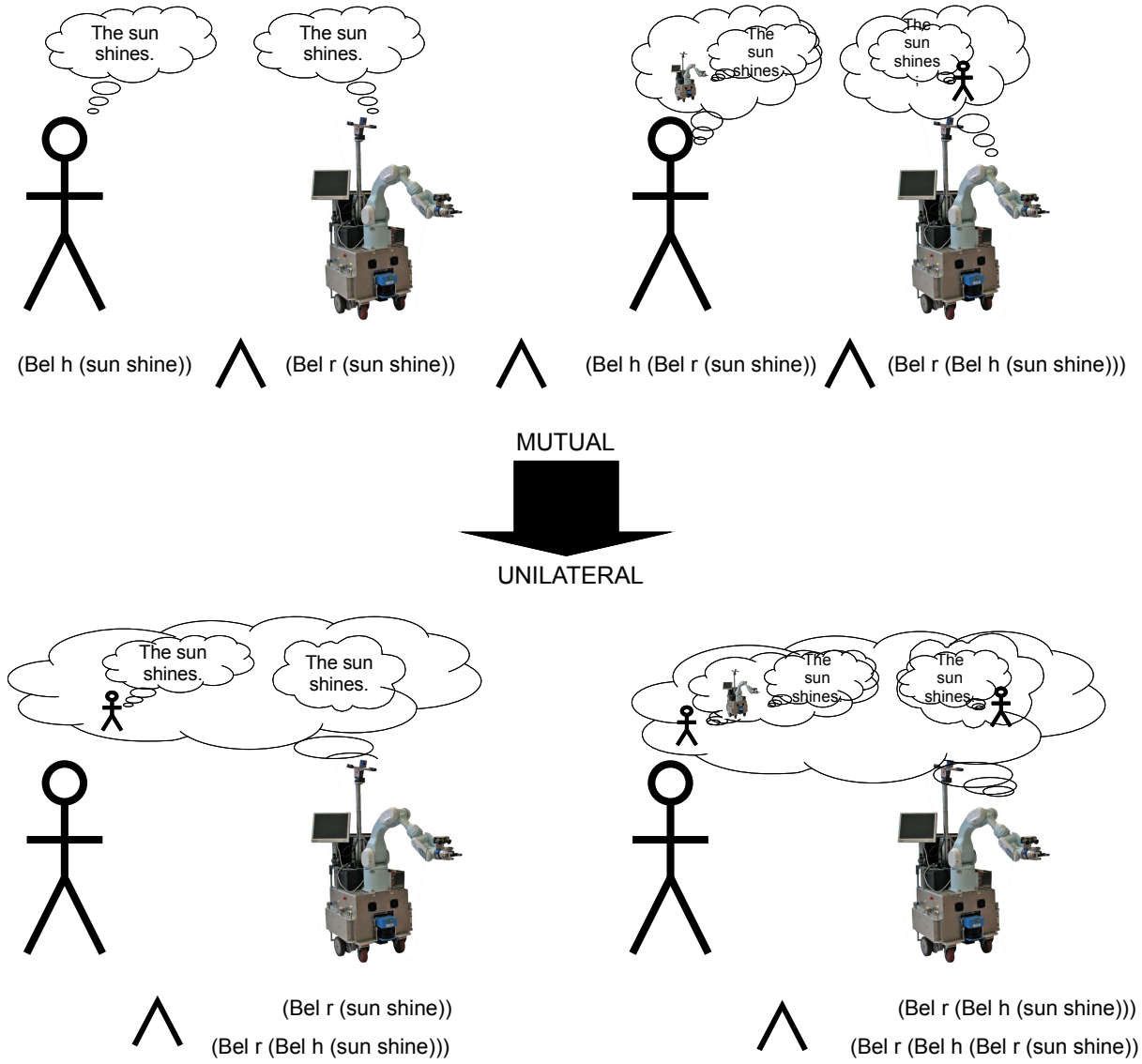


Fig. 2. From Mutual Belief to Unilateral Belief: this scheme represents the translation made from mutual belief to unilateral belief where all beliefs are represented through robot perception.

is that it is useful to give to the visitor an information on the trajectory. Since visitors were not aware of the robot kinematic constraints and path planning algorithms, initial movements were often misunderstood and perceived incoherent. The solution we found consisted in displaying the robot trajectory within the museum map, in order to show that the robot will eventually lead the human to the correct location. The fact that the trajectory is displayed (and the human looks at it) could be interpreted as: $(Bel\ r\ (Bel\ h\ (Bel\ r\ trajectoryknown)))$. If we assume that the human sees the trajectory, we can conclude: $(Bel\ r\ (Bel\ h\ trajectoryknown))$.

Those two examples illustrate the need for the robot to integrate these beliefs at a decisional level. Sometimes, beliefs are even mandatory within the task realization, e.g. when Jido needs to move its arm towards a human, it needs to ensure that they perceive each other. This information is critical for human safety and is therefore a precondition for the arm movement. The robot needs a way to obtain these beliefs in order to complete the task and we will see now how they could be reached.

III. COMMUNICATION

Communication is the act of sharing information and therefore, obtaining shared belief will lead to communication.By communication we mean not only verbal communication but all communication means that can be available (gestures, movements,...). As explain in [15], contribution to communication are not restrained to negative evidence - evidence that we have been misheard or misunderstood, but

also to positive evidence by the way of acknowledgment, relevant next turn or continued attention.

We define a communicative act as an exchange of information between two agents, in our case the robot and the other agents with whom it will interact.A communicative act is defined by a **name** which characterizes the object of the communication (i.e. related to the variable on which we try to obtain mutual belief) and a *step* which characterizes grounding evolution during communication. For example, if we define a communicative act **ask-task**, steps could correspond to the act realisation (**ask-task** *act*), which could be followed by an acknowledgement (**ask-task** *ack*) or an answer (**ask-task** *agree* or *refuse*). We call a sequence of communicative acts a communication scheme that will be defined now.

Our communication schemes embed two mains ideas that are commonly admitted. The first one consists in the fact that each time we communicate we are waiting for something in return. Moreover, we know more or less the set of expected answers or events, and we will use this knowledge to help the system. The second idea is based on the fact that communication implies information exchange; it is not sufficient to consider that an information has been sent, we also need to ensure this information has been correctly understandood by the partner. We introduce also the notion of clarity of the human answer or other communicative act. Clarity can be related to the notion of convention. A convention is a way in which an action is usually executed; it is also defined as a coordination device [19]. The problem for us is the lack of convention existing today between humans and robots, i.e. it is not easy to find the borderline between what could be considered as conventional or not in a human-robot interaction context (not to mention it would be human and context dependent). We will see that an act could be interpreted as non-clear when the robot has an idea of its meaning but is not certain of the true human intention.

Our communication scheme could be compared to Sidner [20] artificial discourse language and Kumar [21] protocols for joint actions, except here we try to consider possible misunderstandings (and their modelings) and adapted procedures to deal with. They could also be linked to Clark levels of grounding [19]: attend, identify, understand and consider.

We will now propose two possible developments through communication scheme of a communicative act. The first one considers the case the robot initiates the communication and the second one it is the human. We'll see how these two developments could be translated through beliefs.

A. Communication R→H

The case where the robot takes the initiative of the communication is described figure 4. These figures show our modeling of the grounding process.

Beliefs are analysed at each step. At the beginning the robot gets its own belief concerning an attribute $(Bel\ r\ (att\ val1))$. By executing the communicative act, it will make the information available to its human partner. It has to be noticed that an information that is not defined as public will never be communicated.

The robot waits at least for an acknowledgment indicating that the human has the information that translates in $(Bel\ r\ (Bel\ h\ (Bel\ r\ (att\ val1))))$.

The robot could also receive what we call a non-clear answer from the human which will be translated in $(Bel\ r\ (Bel\ h\ (att\ val2)))$. This answer could indicate if the human agrees or not with att value, consequently we could have $val1 = val2$ or $val1 \neq val2$.

This translates the fact that the robot thinks the human has a belief $(Bel\ r\ (Bel\ h\ (att\ val2)))$ but it does not know if it is this belief that the human wanted to give, i.e. we do not have (yet) $(Bel\ r\ (Bel\ h\ (Bel\ r\ (Bel\ h\ (att\ val2)))))$ which would translate the fact the human thinks it has given information to the robot that he has the information.

The last proposition consists in the human giving a clear answer, which adds $(Bel\ r\ (Bel\ h\ (Bel\ r\ (Bel\ h\ (att\ val2)))))$.

Beliefs are obtained in the following order :

1) $(Bel\ r\ (att\ val1))$,
2) $(Bel\ r\ (Bel\ h\ (Bel\ r\ (att\ val1))))$,
3) $(Bel\ r\ (Bel\ h\ (att\ val2)))$,
4) $(Bel\ r\ (Bel\ h\ (Bel\ r\ (Bel\ h\ (att\ val2)))))$.

Of course, if beliefs are already present, the entry in the scheme will be adapted.

These four beliefs define $(UMB\ r\ h\ (fact\ val))$ as previously explained.

B. Communication H→R

In the case where the human is the communication instigator, schemes are those of figure 5. In this case the main difference is that before the communicative act occurs, it could be possible that no beliefs are present in the robot knowledge. As detailed before, we have considered the possibility that the robot has difficulties to perceive/understand well the person intentions, that's why we define :

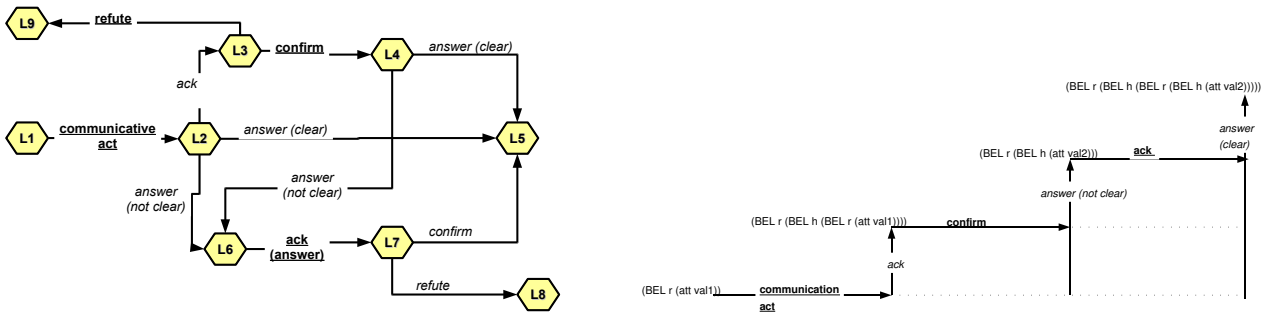
- a *non-clear communicative act*: robot has a belief concerning the human $(Bel\ r\ (Bel\ h\ (att\ val1)))$
- a *clear communicative act*: robot has a belief concerning the human $(Bel\ r\ (Bel\ h\ (att\ val1)))$ and it has the belief that the human knows that $(Bel\ r\ (Bel\ h\ (Bel\ r\ (Bel\ h\ (att\ val1)))))$.

If the communicative act is considered as *non-clear*, the robot sends an acknowledgment to confirm its perception. It waits for a confirmation (or not) to send its answer.

Here also, beliefs are cumulative:

1) $(Bel\ r\ (Bel\ h\ (att\ val1)))$
2) $(Bel\ r\ (Bel\ h\ (Bel\ r\ (Bel\ h\ (att\ val1)))))$
3) $(Bel\ r\ (att\ val2))$
4) $(Bel\ r\ (Bel\ h\ (Bel\ r\ (att\ val2))))$

We observed that the robot needs to take a decision before sending its answer, implying that a dedicated decisional mechanism must be implemented to infer $(Bel\ r\ (att\ val2)))$, i.e. if $val2 = val1$ or $val2 \neq val1$. For instance when the user requires the robot to perform a



Example :
The robot need to share the belief that the human will participate to the task with him.

communicative act: "Do you want to do the task with me ?"
ack communicative act: "You want me to do the task with you ?"
confirm communicative act: "Yes !"
answer clear: "Ok, let's do it !"
answer not clear: Robot detects what it interprets as a head nodding.
ack answer: "It's ok ?"
confirm answer: "Yes !"
refute answer: "oh no !"

Fig. 4. Communication scheme and grounding process evolution when the robot is the communication instigator (given the answer : $val1 = val2$ or $val1 \neq val2$). **underlineandbold** are robot acts and *italic* represent human acts

task, the robot must evaluate its capacity to achieve it in the current context. It will then send an agreement or a refusal.

In the next section we will see how these schemes are developed inside our system.

IV. EXAMPLE

We have defined a set of communicative acts that the human could do at every moment: issue a request, suspend/resume a particular task, modify its plan, etc. At any time, both the user and the robot can propose the following task-based communicative acts (and the corresponding attribute that need to be grounded):

ASK_TASK: proposing a task,
att=task commitment
PROPOSE_TASK_PLAN: proposing a plan (recipe) for a given task,
att= task plan commitment
PROPOSE_MODIFY_PLAN: proposing a modification of the current plan for a given task,
att= task plan commitement
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.
att= (task state=impossible)
CANCEL: cancellation of a task (voluntary give-up),
att= (task state=irrelevant)
TASK_DONE: announces that the task has been done,
att= (task state=achieved))
REALIZE_TASK: announces that the task performance will start. att= (task state=unachieved)

Figure 6 shows an example of the use of these communicative acts concerning a task where the robot has to give an object to a given human : Thierry. To achieve this task, the robot decisional system uses a set of monitors that translates perception into beliefs and acts.

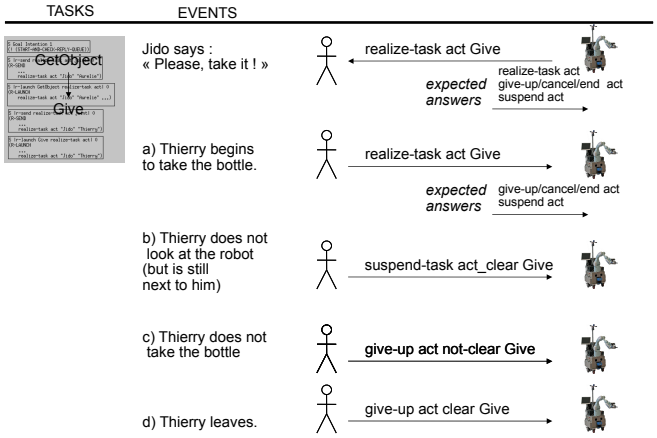
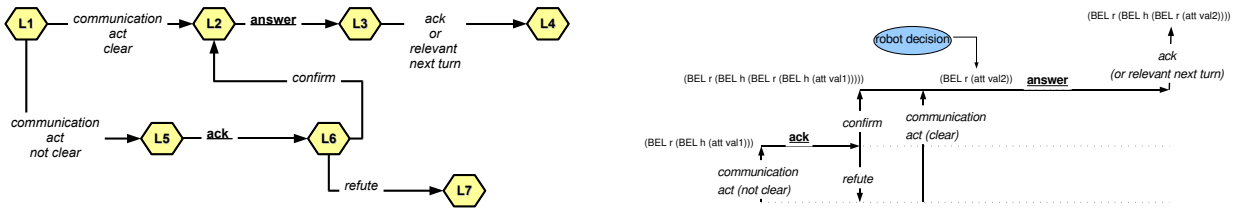


Fig. 6. Example : Jido has to give a bottle to Thierry. First it needs to ensure that Thierry wants to participate to the task, otherwise lack of participation will be interpreted differently given the nature of the detected events. Videos of this task can be found at the following address : http://www.laas.fr/robots/jido



Example :

communicative act clear: "Rackham, we need to suspend the task"
answer: "Ok"
ack answer or relevant next turn: "we agree"
communicative act not clear: The human stops the current collaborative task.
ack communicative act: "Do you want to suspend the task ?"
confirm act: "Yes, we need."
refute act: "No, no, continue !"

Fig. 5. Communication scheme and grounding process evolution when the human is the communication instigator (given the answer : $val1 = val2$ or $val1 \neq val2$). **underlineandbold** are robot acts and *italic* represents human's act

V. CONCLUSION

This work has focused on modeling knowledge-sharing between the robot and the human in the process of interactive task performance. We have modeled and given to the robot not only knowledge concerning itself and the human but also knowledge of the human concerning the robot (i.e. what the human knows or not concerning the robot).

This knowledge helps us to prevent lack of understanding when an information that needs to be shared (i.e. made public) is not. For example, the robot is able to know the human is following him (its perception abilities give the information) but how do we know the human is aware of this robot ability. In order to clarify the situation and maintain shared beliefs between the human and the robot, information need to be sent to the human (e.g. by the help of an external representation displaying the video stream with a box around the detected head or by a gesture or a speech,...).

To help the robot to anticipate what to do given the state of the beliefs, we have defined communication schemes which are dependent on the beliefs coming from the human or the robot. On that basis, we have been able to implement those schemes in the context of human-robot collaborative task achievement on a first set of communicative acts.

REFERENCES

[1] C. Breazeal, "Towards sociable robots," *Robotics and Autonomous Systems*, 2003.

[2] T. W. Fong and Al., "The peer-to-peer human-robot interaction project," *AIAA Space*, 2005.

[3] J. G. Trafton, N. L. Cassimatis, M. D. Bugasjska, D. P. Prock, F. E. Mintz, and A. C. Schultz, "Enabling effective human-robot interaction using perspective-taking in robots," *IEEE Transactions on Systems, Man and Cybernetics*, 2005.

[4] E. A. Sisbot, A. Clodic, L. F. Marin Urias, R. Alami, and T. Siméon, "A mobile robot that performs human acceptable motion," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2006.

[5] R. Alami, A. Clodic, V. Montreuil, E. A. Sisbot, and C. Raja, "Toward human-aware robot task planning," in *AAAI Spring Symposium "To boldly go where no human-robot team has gone before", Stanford, USA*, 2006.

[6] R. Alami, A. Clodic, V. Montreuil, E. A. Sisbot, and R. Chatila, "Towards human-aware cognitive robotics," in *The 5th International Cognitive Robotics Workshop (The AAAI-06 Workshop on Cognitive Robotics), (Cogrob)*, 2006.

[7] T. W. Fong, C. Kunz, L. Hiatt, and M. Bugajska, "The human-robot interaction operating system," in *Human-Robot Interaction Conference (HRI)*, ACM, 2006.

[8] M. Tambe, "Agent architectures for flexible, practical teamwork," *Proc. of the National Conference on Artificial Intelligence (AAAI)*, 1997.

[9] M. Tambe, "Towards flexible teamwork," *Journal of Artificial Intelligence Research*, 1997.

[10] C. Rich, C. L. Sidner, and N. Lesh, "Collagen: Applying collaborative discourse theory to human-computer interaction," *Artificial Intelligence Magazine, Special Issue on Intelligent User Interfaces*, 2001.

[11] C. L. Sidner, "Building spoken language collaborative interface agents," *Technical Report TR2002-038 from Mitsubishi Electric Research Laboratories*, 2002.

[12] R. T. Maheswaran, M. P. Tambe, Varakantham, and K. Myers, "Adjustable autonomy challenges in personal assistant agents: A position paper," *Proceedings of Autonomy'03*, 2003.

[13] P. R. Cohen and H. J. Levesque, "Teamwork," *Nous*, 1991.

[14] P. R. Cohen, H. J. Levesque, and I. Smith, "On team formation," *Contemporary Action Theory*, 1998.

[15] H. H. Clark and S. E. Brennan, *Perspectives on socially shared cognition*. APA Books, 1991, ch. Grounding in communication.

[16] S. Kumar, M. J. Huber, P. R. Cohen, and D. R. McGee, "Toward a formalism for conversation protocols using joint intention theory," *Computational Intelligence*, 2002.

[17] A. Clodic and al., "Rackham: An interactive robot-guide," in *IEEE International Workshop on Robots and Human Interactive Communication (ROMAN)*, 2006.

[18] A. Clodic, S. Fleury, R. Alami, M. Herrb, and R. Chatila, "Supervision and interaction: Analysis from an autonomous tour-guide robot deployment," *12th Int. Conference on Advanced Robotics (ICAR)*, 2005.

[19] H. H. Clark, *Using Language*.Cambridge University Press, 1996.

[20] C. L. Sidner, "An artificial discourse language for collaborative negotiation," in *12th National Conference on Artificial Intelligence*, AAAI Press, 1994.

[21] S. Kumar, M. J. Huber, and P. R. Cohen, "Representing and executing protocols as joint actions," in *1th International Joint Conference on Autonomous Agents & Multi-Agent Systems (AAMAS)*, 2002.