

ACRONYME	ROBOERGOSUM
NOM DU PROJET	ROBOTS CONSCIENTS
REFERENCE	DECISION ANR-12-CORD-0030
NUMERO DE LA TACHE	T5
NOM DE LA TACHE	Self-Awareness and Social Awareness
NUMERO DU RAPPORT	D5.1
TITRE DU RAPPORT	Technical report and submitted paper on supervision of human-robot interaction based on perspective taking
PARTENAIRES	<u>LAAS</u> , ISIR
DATE	T0+18

1. RESUME DU DELIVERABLE / SUMMARY	3
2. CONTRIBUTIONS	4

1. RESUME DU DELIVERABLE / SUMMARY

Dans cette tâche, nous analysons la « conscience » de lui même (self-awareness), des autres et de son environnement (social-awareness) nécessaire au robot pour interagir. L'objectif est de doter le robot des capacités lui permettant de bien comprendre l'interaction et de bien la gérer.

Pour trouver quels étaient ces éléments nécessaires, nous nous sommes penchés sur le domaine de l'action jointe homme-homme. Il est en effet intéressant de constater que de nombreuses recherches sont menées aujourd'hui, aussi bien en psychologie qu'en philosophie, dans ce domaine et qu'il fait l'objet de nombreuses publications. Nous avons pensé qu'il était important de bien comprendre ces recherches.

Nous avons pour cela, mené un premier travail autobiographique, majoritairement axé sur les travaux de Tomasello, Knöblich et Sebanz en psychologie, et Pacherie et Bratman en philosophie. Ce travail intitulé « Key elements pour Human-Robot Joint Action », présent dans la suite de ce deliverable, a fait l'objet d'une présentation à la conférence Robo-Philosophy dont les proceedings sont en cours d'édition dans Frontiers of AI and applications. Une version étendue de ce papier, auquel Elisabeth Pacherie a accepté de participer, est en cours d'écriture et sera édité chez Springer « Studies in the Philosophy of Sociality ».

En analysant l'ensemble de ces travaux, nous avons pensé qu'il était important pour la communauté robotique de s'ouvrir à cette communauté de l'action jointe homme-homme (tant en psychologie qu'en philosophie). Pour permettre cette rencontre, nous avons organisé un workshop pluridisciplinaire à la conférence RO-MAN en août 2014. Ce workshop ainsi qu'un résumé des contributions est présenté dans la suite du livrable. Nous sommes en train d'essayer d'éditer les contributions de ce workshop dans un numéro spécial d'International Journal of Social Robotics. Ce workshop qui a réuni plus d'une trentaine de chercheurs, a permis à des philosophes et des psychologues de présenter les travaux actuels en action jointe homme-homme et aux roboticiens de présentés à ces derniers les défis auxquels ils sont confrontés lorsqu'il s'agit de faire interagir un homme et un robot. Nous présentons également l'exemple qui a servi de fil rouge au workshop. En effet, il nous a semblé important que les chercheurs s'approprient un exemple commun pour permettre un meilleur échange.

Cet exemple a fait l'objet d'une étude et d'une implémentation au LAAS dans le cadre d'un stage de Master réalisé par Sandra Devin. Son rapport de stage est joint à ce livrable.

2. CONTRIBUTIONS

Key Elements for Human-Robot Joint Action

Aurélie CLODIC^{a,1}, Rachid ALAMI^a and Raja CHATILA^{b,c}

^a*CNRS, LAAS, 7 avenue du colonel Roche, F 31400 Toulouse, France and Univ de Toulouse, LAAS, F 31400 Toulouse, France*

^b*Sorbonne Universités, UPMC, Univ Paris 06, UMR 7222, Institut des Systèmes Intelligents et de Robotique, F-75005, Paris, France*

^c*CNRS, UMR 7222, ISIR, F-75005, Paris, France*

Abstract. For more than a decade, the field of human-robot interaction has generated many valuable contributions of interest to the robotics community at large. The field is vast, going all the way from perception to action and decision. In the same time, research on human-human joint action has become a topic of intense research in cognitive psychology and philosophy, bringing elements and even architecture hints to help our understanding of human-human joint action. In this paper, we would like to analyse some findings from these disciplines and connect them to the human-robot joint action case. This work is for us a first step toward the definition of a framework dedicated to human-robot interaction.

Keywords. Action, Joint action, Architecture for Social Robotics, Human Robot Interaction

Introduction

For more than a decade, the field of human-robot interaction has generated many valuable contributions of interest to the robotics community at large. The field is vast, going all the way from perception (e.g., tactile or visual) to action (e.g., manipulation, navigation) and decision (e.g., interaction, human-aware planning). In the same time, research on human-human joint action has become a topic of intense research in cognitive psychology and philosophy, bringing elements and even architecture hints to help our understanding of human-human joint action. In this paper, we would like to analyse some findings from these disciplines and connect them to the human-robot joint action case. More precisely, we are trying to address in this paper the following questions:

- What a robot needs to understand about the human it interacts with for the interaction to be successful and thus what capacities the robot should be equipped with to ensure it can build this understanding?
- On the other hand, the robot also needs to be understood by its human partner. How this understanding operates and what is needed to enable the robot to

¹ Corresponding Author.

behave appropriately and in a way that manifests what it is doing to the human partner

This work is for us a first step toward the definition of an integrative framework needed for the design of an autonomous robot that can engage in interaction with a human partner.

1. Related work and vision

Let's illustrate by a simple example, the kind of interaction we envision. A human and a robot have the goal to build a pile with 4 cubes and put a triangle at the top. There are face to face. One after the other, they should stack bricks in the expected order. Each agent has a number of cubes accessible in front of him and would participate to the task by placing its cubes on the pile. At the end, one of the agents should place a triangle at the top of the pile.

Figure 1 illustrates the initial state. Actions available for each agent are the following (with object = cube or triangle): take an object on the table, take an object from the pile, put an object on the pile, give an object to the other agent, support the pile.

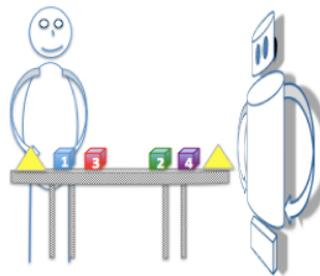


Figure 1: initial state

Each agent is able to infer the state of the world so it knows: where each object is, if an object is reachable for itself, if an object is reachable for the other one. Moreover, we assume each agent is able to observe the activity of the other. Figure 2 presents what the expected final state could be.

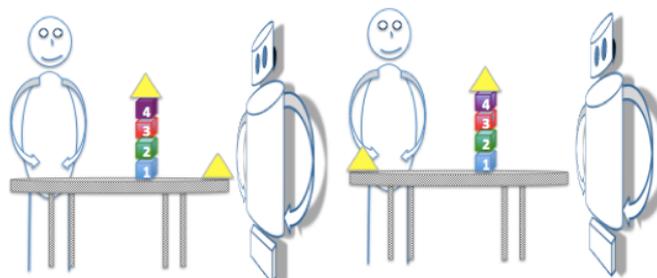


Figure 2: possible final states

Possible deviations could be for example that an agent drops a brick on its side / in the opposite side (e.g. if the brick falls down on the opposite side so that it becomes unreachable for the intended agent to put it on the pile, consider whether the other agent should put the brick directly on the pile or give it to the intended agent) or that the pile collapses. Moreover, during the execution of the task, a number of behaviours can arise, among all: Proactive behaviour (one agent could [be lead to] help the other one by supporting the pile while the other places a brick on it), "Inactive" behaviour (one agent does not act at all) or "Incorrect" behaviour (one agent does not pile bricks in the correct order or one agent removes a correctly placed brick from the pile).

A number of robotics systems, that could be more or less considered as frameworks or architectures dedicated to human-robot interaction have been built, among all [2, 3, 5, 7, 9, 10, 11, 13, 14, 22, 23, 25, 26, 28, 29]. These works have made the robotics community move a step toward understanding human-robot interaction specificities. We want now to have a look on the needed elements to build a frame around all that contributions and this paper is a first step toward this search.

Our aim here is to link human-robot interaction needs to human-human joint action research and to see how it can help to frame such architecture effort.

2. Acting Autonomously

Before entering the joint action domain, we feel necessary to situate the context of autonomous (or individual) action. According to Pacherie [18], today dominant position in philosophical action theory is that "behaviour qualifies as action just in case it has a certain cause or involves a certain sort of psychological process". In the same stream, Tomasello [30] proposes that an "intention is a plan of action the organism chooses and commits itself to in pursuit of a goal. An intention thus includes both a means (action plan) as well as a goal" and that "choosing an intended course of action (decision making), the organism consults both its stored knowledge/skills and its mental model of current reality". We are now equipped, with a definition of an action, an intention, a goal and a plan, elements that should be handled to enable acting.

In the 90's the robotics community tackled the problem of robot control architecture and gave to it several solutions. One of them was the three-layered architecture [12, 1, 16, 17, 25, 29], which defines:

- A functional level which includes all the basic built-in robot action and perception capacities. These processing functions and control loops (image processing, obstacle avoidance, motion control, etc.) are encapsulated into controllable communicating modules. In order to make this level as hardware independent as possible, and hence portable from a robot to another, it is interfaced with the sensors and effectors through a logical robot level. In order to accomplish a task, the next level activates the modules.
- An execution control level, or executive, which controls and coordinates the execution of the functions distributed in the modules according to the task requirements. It is at this level that context-based action refinement is performed.
- A decision level which includes the capacities of producing the task plan and supervising its execution, while being at the same time reactive to events from the previous level. This level may be decomposed into two or more layers,

based on the same conceptual design, but using different representation abstractions or different algorithmic tools, and having different temporal properties.

This architecture relies on representation of action, goal, plan as well as robot's knowledge and skills. However, robot knowledge representation and management is still an open problem.

Interestingly, Pacherie [18][19] proposes an action theory that also distinguishes three main stages in the process of action specification:

- distal intentions level (D-intentions) in charge of the dynamics of decision making, temporal flexibility and high level rational guidance and monitoring of action;
- proximal intentions level (P-intentions) which inherits a plan from the previous level and which role is to anchor this plan in the situation of action, this anchoring has to be performed at two levels: temporal anchoring and situational anchoring;
- motor intentions level (M-intentions), which encodes what neuroscientists call motor representations; with two levels of dynamics: local (specific to each level of intention), global (transition from one level of intention to the next).

This nicely shows a convergence between a philosophical theory of action and a robot control architecture dedicated to action. It seems relevant to have a look if we can build a similar convergence with joint action theory.

3. Acting Jointly

As stated by Knoblich [15] "What distinguishes joint actions from individual actions is that the joint ones involve a shared intention and shared intentions are essential for understanding coordinate joint action". Tomasello [30] says nothing else when he assumes that "Understanding the intentional actions and perception of others is not by itself sufficient to produce humanlike social or cultural activities. Something additional is required. Our hypothesis for this "something additional" is shared intentionality" and more precisely [30] "shared intentionality refers to collaborative interactions in which participants have a shared goal (shared commitment) and coordinated action roles for pursuing that shared goal".

Pacherie proposes a theory of joint action, which also considers three levels of action [20, 21, 22]. If we try to map this theory to robot architecture, we can describe these three levels as the following:

3.1. Shared Distal / Decisional Level

At this level, acting lonely, the robot handles its goal, plan and decision-making; all elements that it represents would be realized by itself. Acting jointly, the robot must be able to handle joint goal, plan and action representation and possibly cooperative decision-making (including e.g. joint planning abilities). It will represent not only what would be achieved by itself but also by the other (with potentially different levels of granularity and completeness). Moreover, high level monitoring would include not only its monitoring but also more generally monitoring of the joint goal and consequently monitoring of the other actions too.

3.2. Shared Proximal / Execution Level

It is at that level that will arise situational and temporal anchoring of the action, which means parameterization of functional level and functions launching and monitoring. At that level, the robot and the human need to be able to share representations (in the best case jointly) and to coordinate their perceptions (to achieve joint attention) in order to coordinate their actions and possibly realize adjustment (dyadic, triadic and collaborative) in the current context.

3.3. Coupled Motor / Functional Level

This level will correspond to robot sensory-motor behaviour that would allow to achieve high-bandwidth interaction with its human partner. An example could be exchanging an object with a human and the associated force-feedback processes. In such tight situation, involving precise coordination between the actors, the parameterization of the functional level needs to be coupled with the one of the other actor. That means, e.g. that the robot control loop would be directly parameterized by the other actor move or action.

We see that this three layers division seems meaningful not only for the human-human case but also for the human-robot case. Having that in mind, we will now explain which elements are needed to setup a framework based on this.

4. What is needed for Joint Action?

We want now to identify and localize the main ingredients and process involved in joint action and how they can make sense in a robotics context. To do so, we will inspire from joint action theory [20], shared intentions theory [30] and other works in psychology of joint action [15] or language [6]; some of those works derive from joint intention theory [8] and shared cooperative activity [4].

According to Knoblich [15]: “a joint action is a social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment”. In [20], Pacherie proposes several dimensions of joint action, in our case, we will consider what she calls: small scale, egalitarian, involving face-to-face interaction. We will first study intentional action understanding as a first step to joint action, then analyse joint attention and elements that need to be shared to end up with a proposal of joint representation definition.

4.1. How to Share?

It is obvious that dialog and negotiation is a way to share [8], but there we will focus on lower level means to achieve joint action.

4.1.1. Intentional Action Understanding

A very interesting prerequisite to joint action established by Tomasello in [30] is understanding of intentional action. We mean that each actor should be able to read its interactor actions. For an observer to understand an intentional action, he must, viewing

an actor's action and more precisely actor's course of actions, be able to: represent the actor's intention (i.e. its goal and plan, and possibly to understand that a choice has been made between several plans) and to understand what the actor is attending to in its perceptive field. This kind of "reverse engineering" process is possible under the assumption that the viewer owns representations about/of the other: its knowledge and skills (and possibly its lack of knowledge) and its model of the current reality as illustrated in the Figure 3 that represents the intentional action understanding. At left, the robot represents the other (in this case the human) and infers what it is doing. At right, the human represents the other (in this case the robot) and infers what it is doing.

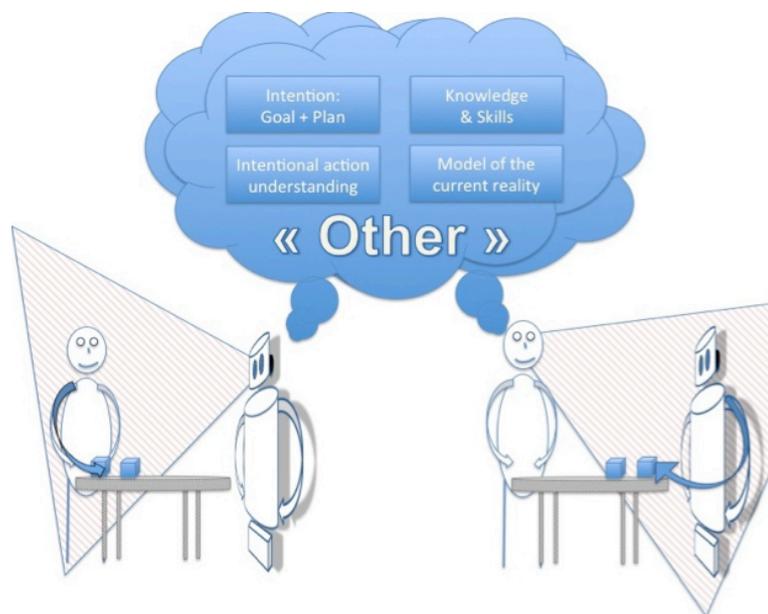


Figure 3: intentional action understanding

What does this say for a robot to understand a human intentional action? That means, we must equip the robot abilities to represent "the other". To this end, the question has to be answered if it could use its proper representations adapted to the "other" or if any other form of representations is needed. This capability should be, of course, limited to the context and the tasks the robot will be involved into: for instance navigation and associated activities, or simple object manipulation in domestic environment.

On the other side, for a human to understand a robot intentional action, he must have access to robot knowledge and skills, this means the robot should be (and behave so that to be) understandable to the human. That means too, that the human must be able to infer the robot model of the current reality and it is not so simple since the robot sensing abilities are not fully readable by the human.

4.1.2. Joint Attention

One key means to share perceptual representation in face-to-face interaction is joint attention. Attention is the cognitive process of selectively concentrating on one aspect

of the environment while ignoring other things (from <http://en.wikipedia.org/wiki/Attention>). Pacherie defines joint attention as “two people attending to the same object or event + actual attention sharing (there must be some causal connection between the two subjects’ acts of attending) + mutual manifestness (the fact that both are attending to the same object or event should be open or mutually manifest)”([20] page 355). This concept, that we could find too in [30] or [15] for example, is key because it states that if joint attention is established, whatever information I can get, I can consider my interactor would have it too if it occurs in the joint attention space. It includes what both interactors perceive, but also what only one interactor perceives (e.g. if one part of the table is hidden to the robot, the robot can establish that it cannot see a part of the environment, whereas the human is able to see this part - and vice-versa, the robot can assume the human knows that a part of the table is hidden to the robot and that the human can see this part.). This raises a number of questions: How can a robot know that the human it interacts with joint attended with him to the joint task? What are the cues that should be collect to infer joint attention? Symmetrically, how can a robot exhibit joint attention? What cues the robot should exhibit to let the human infer that joint attention is met? Moreover, once joint attention is achieved (or at least a given level of joint attention if we consider it is not a 0/1 option), how should it be managed during the overall course of joint actions? Tomasello [30] explains that actors need to handle cooperative perception while joint goal unfolds. How can we handle cooperative perception? Does this need to be taken into account at planning level or at anchoring level?

Under joint attention assumption, in the joint attention space, all events that happen are supposed to be shared between the interactors. It has to be noticed that this information needs to be filtered by perspective taking abilities ([15]): if the robot is in face of the human and perceives that brick 1 is at its left side, it should infer that brick 1 is at the right side of human. Moreover, that does not say anything about what both the interactors perceive means for both of them. It is there shared action/task/goal representation is helpful.

4.2. What to Share ?

In [20], Pacherie establishes that a number of elements must be handled by each agent to drive a joint action:

- self-predictions: agents each represent their own actions and their predicted consequences in the situation at hand.
- other-predictions: agents each represent the actions, goal, motor and proximal intentions of their coagents and their consequences.
- dyadic adjustment: agents each represent how what they are doing affects what others are doing and vice-versa and adjust their actions accordingly.
- joint action plan: agents each have a representation (which may be only partial) of the hierarchy of situated goals and desired states culminating in the overall joint goal
- joint predictions: agents each predict the joint effects of their own and other’s actions
- triadic adjustment: agents each use joint predictions to monitor progress toward the joint goal and decide on their next moves, including moves that may involve helping others achieve their contributions to the joint goal

In our context that could be illustrated by Figure 4.

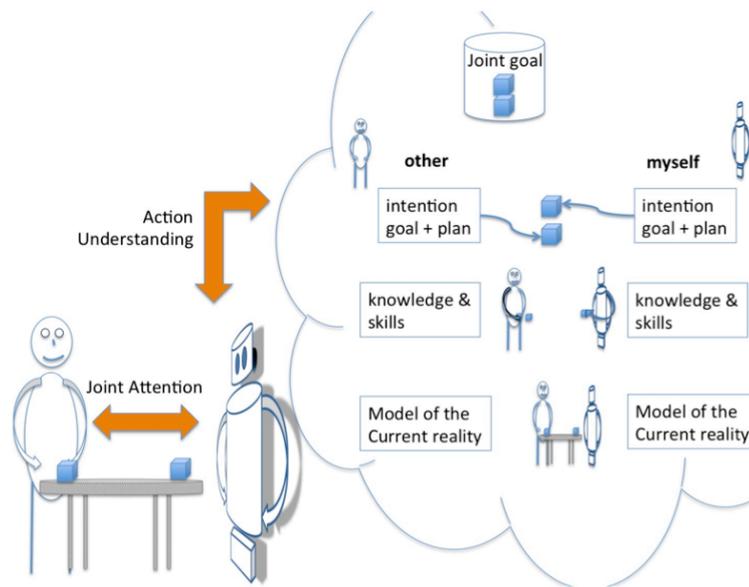


Figure 4: What to Share ?

That means the robot needs to be able to handle: its world representation, a world representation of the human it interacts with (again potentially limited to the task to perform), the possible effect of its actions on the human actions (and vice-versa), their joint goal and action plan representation, a prediction of their actions, a mean to monitor progress toward the joint goal (and possibly mean to revise the on-going joint plan). A triadic adjustment means that the robot and the human can adapt their behaviour toward the joint goal. That means for example, that if the human brings down its brick in the robot space, the robot will place the brick on the stack. If it had done a dyadic adjustment it had make accessible the brick to the human to let him finish its action, a dyadic adjustment means that the robot and the human can adapt their behaviour to the other actions (not toward the joint goal). It has to be noticed that Tomasello [30] does not use exactly the same nomenclature and adds another adjustment (engagement in its vocabulary) level: the collaborative level where he considers the two must plan together toward the joint goal (he does not consider that it is done at previous levels of dyadic and triadic) that could handle behaviour where the human can hold the stack while the robot places its last brick (from [30] page 682). To be able to deal with such elements, the robot must share representations with the human it interacts with: perceptual representation, (joint) action/task/goal representation. This idea of sharing representation drifts from shared intentionality [30], shared intention [15] or interdependence of the individual intentions [21]. However, it has to be noticed that shared representation does not mean common representation. Representations could differ, the important thing is that we are aware of.

4.3. Where to Share ?

Representation sharing could be helped by mechanism such as affordance. From [15], object affordances are the action opportunities that an object provides for an agent with a particular action repertoire whereas common affordance states that when two agents have similar action repertoires and perceive the same object, they are likely to engage in similar actions because the object affords the same action for both of them. That means we must give to the robot access to object and common affordances model to help its human understanding.

Representation sharing would help the robot to achieve perception-action matching [20] in 2 directions: action-to-goal prediction (goal attribution to observed action execution) and goal-to-action prediction (anticipate the observed actor's next actions). It goes in the same direction as [15] common predictive models: "action simulation can lead to emergent coordination because it induces the same expectations about the unfolding of actions in different actors and thus induces similar action tendencies for future actions". Consequently that would help actions prediction and monitoring and also enable dyadic, triadic and collaborative adjustment.

That means that:

- perceiving an object, the robot and the human it interacts with must share information such as:
 - I perceive the object, you perceive the object, I know you know what I perceive;
 - I know what it is (or not), I know you know what it is (or not), I know you know what I know;
 - I know what is its purpose (or not), I know you know its purpose (or not), I know you know what I know;
 - I know how to handle the object (or not), I know you know how to handle the object (or not), I know you know what I know;

e.g. for a robot seeing a telephone on the table next to the human, what would be inferred/shared?

- perceiving an action, the robot and the human it interacts with must share information such as:
 - I perceive the action, you perceive the action, I know you know what I perceive;
 - I know what it means (or not), I know you know what it means (or not), I know you know what I know;

e.g. for a robot, viewing the human scratching its head, what would be inferred/shared? The needed information, its various levels and the way to represent it, even it has been studied in different ways remains an open question that need to be tackle. Moreover, in case of loss or lack of information sharing, the robot needs to be able to inform or facilitate the information acquisition of the human. This has to do with expressive multi-modal behaviour and what is called mutual manifestness (see below).

In addition, we have the intuition that shared representation seems not enough in the context of human-robot interaction. We need what we can call joint representation, i.e. if we refer to Pacherie joint attention definition, we could define joint representation as: "two people sharing a representation + actual representation sharing (there must be some causal connection between the two subjects' acts of sharing a representation) + mutual manifestness (the fact that both are sharing the representation should be open or mutually manifest)". This means that we know (or not) a representation should be

shared mutually. In human-human interaction, assumption can be easily made from both sides on what the other knows or not, this is far more difficult in human-robot interaction. On the robot side, this indicates that we need to integrate into the robot means to share representations explicitly with the human but also means to recognize and understand them (and to learn them if needed). On the other side, a human interacting with a robot is often disconcerted because it is difficult for him to have intuitions about robot capabilities and inabilities. What is missing here is what Tomasello [30] named cultural creation/learning, or Clark [6] common ground, and that is something we need to come up with.

5. Conclusion

In this paper we have proposed an analysis of some findings in Psychology and Philosophy in the domain of human-human joint action in order to come up with needs in terms of knowledge and abilities that a robot, interacting with a human, need to handle.

We propose that intentional action understanding, joint attention and joint representation management are key elements to better human-robot interaction unfolding. Then, we've seen that framework proposed by Philosophy such as [22] could be inspiring in the search to frame an architecture dedicated to human-robot interaction.

This work is a first step toward the objective to identify and incrementally give an accurate description of the different needed abilities and how they are involved in the overall process of collaborative human-robot task achievement.

Future steps would be to continue to analyse inputs from Philosophy and Psychology and to analyse if our requirements have been already implemented in a robotics architecture (even part of) and how. From this basis, we will continue to try to formalize (when possible) and to devise the pertinent human and task related models and the associated decision-making, planning and situation assessment processes.

6. Acknowledgments

This work was conducted within ANR-CONTINT ROBOERGOSUM project (DECISION ANR-12- CORD-0030-02)

References

- [1] R. Alami, R. Chatila, S. Fleury, M. Ghallab, F. Ingrand, An architecture for autonomy, *International Journal of Robotic Research* (1998)
- [2] C. Breazeal, Towards sociable robots, *Robotics and Autonomous Systems* (2003)
- [3] C. Breazeal, M. Berlin, A. Brooks, J. Gray, A. Thomaz, Using perspective taking to learn from ambiguous demonstrations, *Robotics and Autonomous Systems* (2006)
- [4] M. E. Bratman, Shared cooperative activity, *The Philosophical Review* (1992)
- [5] A. Cangelosi, Grounding language in action and perception: From cognitive agents to humanoid robots, *Physics of Life Reviews*, 7(2), 139-151 (2010)
- [6] H. H. Clark, *Using Language*, Cambridge University Press, 1996.

- [7] A. Clodic, H. Cao, S. Alili, V. Montreuil, R. Alami, R. Chatila, Shary: A supervision system adapted to human-robot interaction, Springer Tracts in Advanced Robotics (2009)
- [8] P. R. Cohen, H. J. Levesque, Teamwork, *Nous* (1991)
- [9] Y. Demiris, Prediction of intent in robotics and multi-agent systems, *Cognitive Processing*, 8: 151-158 (2007)
- [10] F. Ferland, D. Létourneau, A. Aumon, J. Frémy, M.A. Legault, M. Lauria, F. Michaud, Natural interaction design of a humanoid robot, *Journal of Human-Robot Interaction* (2012)
- [11] T. W. Fong, C. Kunz, L. Hiatt, M. Bugajska, The human-robot interaction operating system, Proc. Conference on Human-Robot Interaction (HRI) (2006)
- [12] E. Gat, Integrating planning and reacting in a heterogeneous asynchronous architecture for controlling real-world mobile robots, in Proceedings of the tenth national conference on Artificial intelligence, AAAI Press, pp. 809–815 (1992)
- [13] B. J. Grosz, S. Kraus, Collaborative plans for complex group action. *Artificial Intelligence* (1996)
- [14] M. Johnson, Y. Demiris, Perceptual perspective taking and action recognition, *Advanced Robotic Systems* (2005)
- [15] G. Knoblich, S. Butterfill, Psychological research on joint action: theory and data. In B. Ross (Ed.), *The Psychology of Learning and Motivation*, 2011.
- [16] N. Muscettola, P. P. Nayak, B. Pell, and B. C. Williams, Remote agent: To boldly go where no ai system has gone before, *Artificial Intelligence* (1998)
- [17] I. Nesnas, A. Wright, M. Bajracharya, R. Simmons, T. Estlin, CLARAty and challenges of developing interoperable robotic software, IEEE/RSJ Internat. Conf. on Intell. Robots and Systems (IROS) (2003)
- [18] E. Pacherie, The phenomenology of action: A conceptual framework, *Cognition* (2008)
- [19] E. Pacherie, Action, In K. Frankish & w. Ramsey (eds.), *The Cambridge handbook of cognitive science*. Cambridge University Press. 2012
- [20] E. Pacherie, The phenomenology of joint action: Self-agency vs joint-agency. In Axel Seemann (ed.), *Joint Attention: New Developments*, Cambridge MA: MIT Press, 2012
- [21] E. Pacherie, Is collective intentionality really primitive? In M. Beaney, C. Penco & M. Vignolo (Eds.), *Mental processes: representing and inferring*, Cambridge, Cambridge Scholars press (2007)
- [22] E. Pacherie, Framing joint action, *Review of Philosophy and Psychology* (2011)
- [23] A. K. Pandey, A.K., Alami, R.: Towards effect-based autonomous understanding of task semantics for human-robot interaction. *International Journal of Social Robotics (IJSR)* (2013)
- [24] M. Petit, S. Lallée, J. Boucher, G. Pointeau, P. Cheminade, D. Ognibene, E. Chinellato, U. Pattacini, I. Gori, U. Martinez-Hernandez, H. Barron-Gonzalez, M. Inderbitzin, A. Luvizotto, V. Vouloutsi, Y. Demiris, G. Metta and P. Dominey, The Coordinating Role of Language in Real-Time Multi-Modal Learning of Cooperative Tasks, *IEEE Transactions on Autonomous Mental Development*, 5:1, pp 3-17, (2013)
- [25] G. N. Saridis, Architectures for intelligent controls, *Intelligent control systems: Theory and applications*, IEEE Press, (1995)
- [26] M. Scheutz M., Computational Mechanisms for Mental Models in Human-Robot Interaction, *HCI International*, 304-312, (2013)

- [27] C. L. Sidner, C. Lee, C. Kidd, N. Lesh, C. Rich, Explorations in engagement for humans and robots, *Artificial Intelligence* (2005)
- [28] E.A. Sisbot, A. Clodic, R. Alami, M. Ransan, Supervision and motion planning for a mobile manipulator interacting with humans, 3rd ACM/IEEE International Conference on Human-Robot Interaction (2008)
- [29] M. Tambe, Towards flexible teamwork, *Journal of Artificial Intelligence Research* (1997)
- [30] M. Tomasello, M. Carpenter, J. Call, T. Behne, H. Moll, Understanding and sharing intentions: The origins of cultural cognition, *Behavioral and Brain Sciences* (2005)
- [31] J. Trafton, N. Cassimatis, M. Bugajska, D. Brock, F. Mintz, A. Schultz, Enabling effective human-robot interaction using perspective-taking in robots, *IEEE Transactions on Systems, Man, and Cybernetics* (2005)



towards a Framework for Joint Action

fja2014 at RO-MAN

25 Aug 2014 Edinburgh (United Kingdom)

PRESENTATION

The [RO-MAN 2014](#) Workshop "Towards a Framework for Joint Action" (FJA2014) is a full-day workshop held in conjunction with the 23rd IEEE International Symposium on Robot And Human Interactive Communication ([IEEE RO-MAN 2014](#)), in Edinburgh on August 25 2014. This workshop aims to bring together researchers from several disciplines to discuss the development of frameworks for thinking about and designing human-robot joint action.

For more than a decade, the field of human-robot interaction has generated many valuable contributions of interest to the robotics community at large. The field is vast, going all the way from perception (e.g., tactile or visual) to action (e.g., manipulation, navigation) and decision (e.g., interaction, human-aware planning). However, when it comes to the development of future robot assistants or robotic team-mates in mixed human-robot teams, there is a need for a deeper understanding of human-robot joint action that could provide a framework for the different contributions and studies.

It is interesting to observe, from a roboticist point of view, that research on human joint action is a topic of intense research in cognitive psychology and philosophy. In this workshop, we would like to analyse the fundamental assumptions as well as detailed empirical findings from these disciplines and connect them with various ongoing research activities in robotics, from the design of control architectures to human-robot interaction.

More specifically, our goal is to bring various aspects of existing work together and examine how they can help us define the kind of integrative framework needed for the design of an autonomous robot that can engage in long-term interaction with a human partner. This framework should be able to serve two complementary purposes. On the onehand, it should help us define with precision what a robot needs to understand about the human it interacts with for the interaction to be successful and thus what capacities the robot should be equipped with to ensure it can build this understanding. On the other hand, the robot also needs to be understood by its human partner and this framework should help us clarify how this understanding operates and what is needed to enable the robot to behave appropriately and in a way that manifests what it is doing to the human partner.

The aim of this workshop is to bring together researchers in robotics, psychologists and philosophers. This will create a unique opportunity for scientific exchange between these disciplines. In particular, psychologists and philosophers will be given the opportunity to present recent developments in joint action research, while roboticists will be able to discuss the challenges they face with regard to human-robot interaction and more precisely human-robot joint activity. We imagine and hope that the workshop

will be a first meeting in a series of workshops in the coming years that will create a context in which interested researchers can come together and discuss the development of an integrative framework for research on human-robot joint action.

The workshop will be divided in two parts. In the morning, speakers will provide an overview of recent work on joint action from philosophical and psychological perspectives, followed by an overview of current challenges in robotics regarding human-robot and robot-robot joint action.

In the afternoon, participants will be invited to present their own work and ideas regarding key issues in joint action research in the different fields. This session will be followed by a panel discussion with the goal of identifying key elements for establishing robust joint action performance in robotics and ideally taking steps towards the general framework we are seeking.

Speakers

Rachid Alami - CNRS senior researcher at Laboratoire d'Analyse et d'Architecture des Systèmes LAAS-CNRS, Toulouse, France

Elisabeth Pacherie – CNRS senior researcher in Philosophy at Institut Jean Nicod (UMR 8129, ENS, EHESS, CNRS), affiliated with the Institute for the Study of Cognition at Ecole Normale Supérieure, Paris

Julie Shah – Assistant Professor in the Department of Aeronautics and Astronautics, lead the Interactive Robotics Group in CSAIL

Cordula Vesper – Social Mind and Body group (SOMBY), Department of Cognitive Science, Central European University, Budapest, Hungary

CALL FOR PAPERS

Aim

The aim of this workshop is to allow researchers interested in joint action, roboticists but also philosophers and psychologists, to have a context for discussion about and progressive elaboration of a framework for human-robot joint action.

To achieve this goal, we propose to the community to tackle a common example (as it is sometimes done in robotics planning competition) with the goal to identify the capacities and skills needed for the successful performance of the joint action and to see which of these are present or missing in any of our architectures. This should enable us to build upon each other's experience to further develop ongoing work. We hope that this could be a first meeting in a series of workshop in the next years.

Topics

We are seeking to frame joint action, interesting topics include:

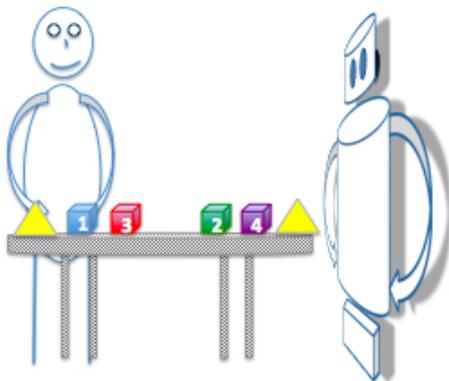
- joint goal establishment and negotiation
- planning when the goal is a joint goal
- joint goal / joint action description and modeling
- joint goal / joint action execution, monitoring, turn-taking and timing management
- agents world modeling and management (incl. theory of mind and shared knowledge)
- agents commitment
- what kind of communication during joint goal achievement and for what purpose

An illustrative example

Please feel free to play with this example since discussion is really welcome. However, it would be very useful if you clearly indicate in what way your work is helpful to deal with such an example, i.e. which "brick" your work could contribute to the framework.

A human and a robot have the goal to build a pile with 4 cubes and put a triangle at the top. One after the other, they should stack bricks in the expected order. Each agent has a number of cubes accessible in front of him and would participate to the task by placing its cubes on the pile. At the end, one of the agent should place a triangle at the top of the pile.

The initial state is the following:



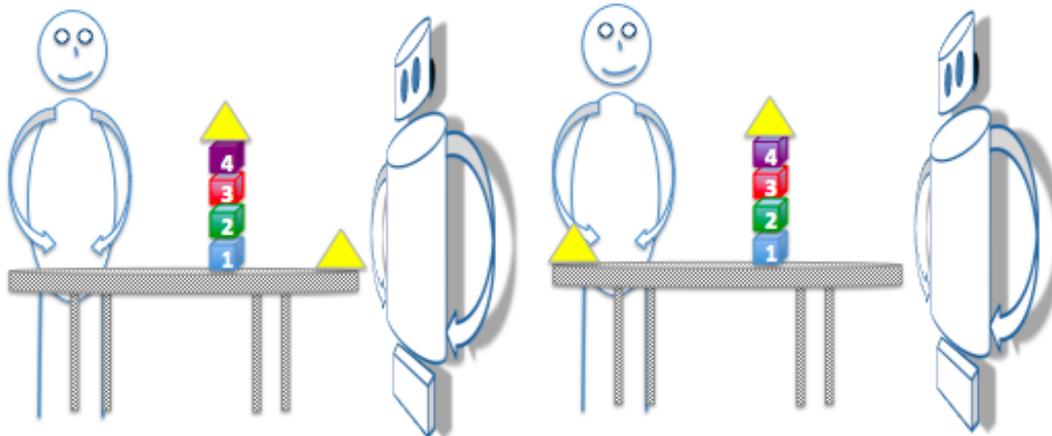
Actions available for each agent are the following (with object = cube or triangle):

- take an object on the table
- take an object from the pile
- put an object on the pile
- give an object to the other agent
- support the pile

Each agent is able to infer the state of the world so it knows:

- where each object is
- if an object is reachable for itself
- if an object is reachable for the other one

Moreover, we assume each agent is able to observe the activity of the other. The expected final state could be one of the following:



Possible deviations could be that:

- an agent drops a brick on its side / in the opposite side
 - e.g. if the brick falls down on the opposite side so that it becomes unreachable for the intended agent to put it on the pile, consider whether the other agent should put the brick directly on the pile or give it to the intended agent
 - the pile collapses
 - ...

Moreover, during the execution of the task, a number of behaviors can arise, among all:

- Proactive behavior:
 - one agent could [be lead to] help the other one by supporting the pile while the other places a brick on it
 - "Inactive" behavior:
 - one agent does not act at all
 - "Incorrect" behavior:
 - one agent does not pile bricks in the correct order
 - one agent removes a correctly placed brick from the pile

Finally, a negotiation phase should arise at the end to decide who put the triangle on the pile.

Topics

We are seeking to frame joint action, interesting topics include:

- joint goal establishment and negotiation
- planning when the goal is a joint goal
- joint goal / joint action description and modeling
- joint goal / joint action execution, monitoring, turn-taking and timing management
- agents world modeling and management (incl. theory of mind and shared knowledge)
- agents commitment
- what kind of communication during joint goal achievement and for what purpose

Application format

We are seeking several kinds of contribution, so you can whether:

- propose a description of an existing system and explain how it can be used in this context
- propose a description of an element, a brick that could be used in (a part of) this context
- propose a position paper where you explain from your point of view, what are the challenges and key elements that should be handle to solve this problem

We invite submission of short papers (max 4 pages) but any other form of submission is welcome in this first phase, too, since we are interested in any contributions that could stimulate a fruitful debate.

PROGRAM

9:00 – 9:10 Welcome

Psychological and philosophical foundations of joint action

9:10 – 9:50

Cognitive Mechanisms Supporting Human Joint Action

Cordula Vesper

Department of Cognitive Science, Central European University, Budapest, Hungary

How do people perform actions together such as lifting a heavy box, playing ensemble music or building a complex object (such as a brick tower) together? A variety of recent empirical studies in psychology, cognitive science and neuroscience shows that planning and performing joint actions is based on specific cognitive mechanisms that support coordination with other people. These include representing others' actions, predicting and monitoring what others do and adapting own actions in ways that facilitate interpersonal coordination. In this talk, I will discuss these mechanisms as well as supporting evidence to provide a short overview about the state-of-the-art of human joint action in order to identify links to human-robot interaction.

-
9:50 – 10:30

Levels of coordination in joint action

Elisabeth Pacherie

Institut Jean Nicod, CNRS-EHESS-ENS, Paris, France

I describe and motivate a dynamic hierarchical model of intention and action specification and use it to propose an analysis of the different levels of coordination involved in joint action and of the means at our disposal to achieve coordination.

-
10:30 – 11:00 Coffee break

Joint action in robotics

-
11:00-11:30

On Human-Aware Planning Abilities for a teammate robot

Rachid Alami

CNRS-LAAS, Université de Toulouse, Toulouse, France

In this talk, I will consider several key decisional issues that are necessary for a cognitive robot which shares space and tasks with a human. We have adopted a constructive approach based on the identification and the effective implementation of a set of decisional processes. I will first give a broad view of a conceptual robot control architecture specially designed to provide a framework for these decisional processes. These abilities include geometric reasoning and situation assessment based essentially on perspective-taking and affordances, management and exploitation of each agent (human and robot) knowledge in a separate cognitive model, human-aware task and motion planning and human and robot interleaved plan achievement.

-
11:30-12:00

Human-Robot Team Training

Julie Shah

Massachusetts Institute of Technology

In this talk, I discuss the design of new models for robot planning, which use insights and data derived from the planning and execution strategies employed by successful human teams, to support more seamless robot participation in human work practices. This includes models for human-robot team training, which involves hands-on practice to clarify sequencing and timing of actions, and for team planning, which includes communication to negotiate and clarify allocation and sequencing of work. The aim is to support both the human and robot workers in co-developing a common understanding of task responsibilities and information requirements, to produce more effective human-robot partnerships.

-
12:00-12:30

Discussion

12:30 - 13:30 Lunch break

Contributions

-
13:30-13:45

(How) Can Robots Make Commitments ? -- A pragmatic Approach

John Michael (Central European University (CEU)), *Alessandro Salice* (Copenhagen University (KU))

Commitment is a fundamental building block of social reality. In particular, commitments seem to play a fundamental role in human social interaction. In this paper, we discuss the possibility of designing robots that engage in commitments, are motivated to honor commitments, and expect others also to be so motivated. We identify several challenges that such a project would likely confront, and consider possibilities for meeting these challenges.

-
13:45-14:00

Human Robot Collaboration in production environments

Christoph Strassmair (Heriot-Watt University (UNITED KINGDOM) and Technische Hochschule Ingolstadt (THI)), *Nick Taylor* (Technische Hochschule Ingolstadt (THI))

Human Robot Collaboration has great potential for the manufacturing domain. A necessary precondition for this type of Joint Action is however the human's acceptance of their robotic co-worker. Recent research indicates that this acceptance is heavily influenced by the flexibility granted to the worker and the efficiency of the collaboration. Current systems aim for efficiency but neglect the spatial constraints of the collaboration. Furthermore they restrict the worker's flexibility.

This paper presents an approach that extends prior art by considering spatial constraints and granting more flexibility to the worker. It thus facilitates a more efficient collaboration and a higher worker acceptance in manufacturing environments. The application of the approach to the workshop example is sketched.

-
14:00-14:15

Role Distribution in Synchronous Human-Robot Joint Action

Tariq Iqbal, *Laurel Riek* (Department of Computer Science and Engineering, University of Notre Dame)

Robots are becoming a part of our daily lives, and humans and robots are beginning to work together as teams to achieve common goals. In a human-robot interaction (HRI) scenario, it is important to assign roles to both the human and the robot, as these role

assignments may affect the fluidity and the effectiveness of the interaction. In this paper, we discuss different role distribution models to assign roles among humans and robots in the context of synchronous joint action. We employed the leader-follower model in a human-robot collaborative task using a method from our previous work to detect synchronous actions of team members. Our results support our choice of the leader-follower model, and suggest that our method is capable of measuring synchronous joint action in an HRI scenario. These results are encouraging for future work aimed at the development of adept human-robot teamwork.

-
14:15-14:30

Joint action and joint attention: what could be the interface between developmental psychology and robotics to better frame human-robot joint activity

Michèle Guidetti (Unité de Recherche Interdisciplinaire Octogone (EA 4156) (URI Octogone), Université Toulouse le Mirail - Toulouse II)

The aim of the proposed talk is to present the point of view on joint attention and joint action of a developmental psychologist interested by normal pathological comparisons and believing that “development is the key to understanding developmental disorders” (Karmiloff-Smith, 1998), knowing nothing about robots but interested by exchanges and collaborations with roboticists in order to better frame human-robot joint activity.

-
14:30-14:45

Language Use in Joint Action: Demonstratives and Nouns used in Japanese Conversation Doing a Collaborative Task

Harumi Kobayashi (Tokyo Denki university), *Tetsuya Yasuda* (Jumonji University), *Hiroshi Igarashi*, *Satoshi Suzuki* (Tokyo Denki university)

English demonstratives “this” “that” “here” “there” are used to quickly and exactly establish joint attention on some objects in the environment with other people. People may say, “Look at this,” or “I want to ride on that bus,” to jointly look at the same object with one or more people. Interestingly, in Japanese, there are no articles, so it is completely grammatical to say nouns without articles. If someone wants to establish joint attention on a specific clock, the person must use a demonstrative such as “Kono tokei wo kaitai (I want to buy this clock)”. In this study, we focused on the use of Japanese demonstratives such as here (in Japanese, “koko”), there (“asoko”), this (“kore,” “kono”), that (“sore,” “sono,” “are,” “ano”) using corpus data and analyzed conversation to examine whether the use of these words change when people execute a collaborative task ten times. We used virtual space using computer monitors. There were three adult volunteers in each group and, there were four such groups. The task was to collaboratively move boxes to predetermined target panels in a virtual space. We extracted demonstratives and nouns and examined the use of these words over ten trials. The results were that the ratio of nouns decreased whereas the ratio of demonstratives increased in three of four groups' data. People may regulate joint actions by effective use of demonstratives when they cooperate with other people in addition to actions. The study also suggested that demonstrative “this” might be a powerful tool to show ones own will to execute the next task. By using the demonstrative “THIS,” robots and humans may be able to quickly communicate who will move the next block in the workshop example.

-
14:45-15:00

Movement coordination in repetitive joint action tasks: Considerations on human-human and human-robot interaction

Tamara Lorenz (Technische Universität München (TUM), Ludwig-Maximilians-Universität [München] (LMU)), *Sandra Hirche* (Technische Universität München (TUM))

It is a common idea that robotic design for human-robot interaction can benefit from approaches taken from human joint action research. In this position paper a pre-defined example is operationalized to shed light on recent findings in human movement coordination and turn-taking in repetitive tasks. Both topics are also considered for human-robot interaction. The paper closes with a discussion on open questions and unsolved problems that are not considered in human psychological research but play a major role when a transfer to robotics is intended.

15:00 – 15:30 Coffee break

15:30-15:45

Using Virtual Characters in the Study of Mimicry and Joint Action

Xueni Pan, Harry Farmer, Joanna Hale, Antonia Hamilton (Institute of Cognitive Neuroscience, UCL)

In recent years, the use of Virtual Characters in experimental studies has opened new research venues in social psychology and neuroscience. In this paper, we review the literature on the use of Virtual Reality in the study of joint action and mimicry, and then present the design and implementation of three case studies in this area using Virtual Characters. Our preliminary results supported the hypothesis that the congruent and mimicry effect exists in the interaction between human-participants and Virtual Characters. Finally, we discuss different types of co-actions as well as how the benefit of using Virtual Reality in this area.

15:45-16:00

Exploring Joint Action for Alternative Finding: Proposal of a Human-Human Study to Inform Human-Robot Collaboration

Astrid Weiss, Markus Vincze (Automation and Control Institute, Vienna University Of Technology (ACIN, TUW))

This paper proposes a study in order to explore how human collaborators monitor the perceptual regions in unsolvable task situations (e.g. the right piece cannot be found in order to build the tower together) and how the collaborators (in our case a director and a builder) use the information in the course of collaboration. The interesting aspect of this study is to see which grounding techniques participants use to identify that the task is unsolvable (i.e. needs an alternative solution) and which of these techniques can be transferred to Human-Robot Collaboration. We propose human-robot grounding performed by the human, meaning to enable the robot to “explain itself” in unsolvable task situations, therefore that the human can establish a common ground and find an informed alternative solution. In this position paper we present the study design and preliminary results, as we are currently in the middle of the data analysis.

16:00-17:00

General Discussion

Discussion involving all speakers and workshop attendants

Laboratoire d'Analyse et d'Architecture des Systèmes (LAAS-CNRS),
Toulouse

Projet de fin d'étude

Maître de stage : Rachid Alami

Vers l'action jointe homme-robot

Sandra DEVIN



Remerciements

Je tiens tout d'abord à remercier Rachid ALAMI de m'avoir permis de réaliser ce stage.

Je le remercie aussi ainsi qu'Aurélie CLODIC pour leurs encadrements et les conseils qu'ils m'ont apporté.

Merci à toute l'équipe de doctorants et stagiaires pour leurs conseils, leur accueil et la bonne ambiance constante tout au long du stage.

Enfin, merci au robot PR2 Max pour sa patience durant ces longues heures d'implémentation.

Table des matières

1	Introduction	1
1.1	Présentation du problème	2
2	Études préliminaires	3
2.1	État de l'art sur l'action jointe	3
2.1.1	Théorie de l'action jointe, ce qui est nécessaire	3
2.1.1.1	Le partage du but et de l'intention	5
2.1.1.2	La perception et la prédiction des actions de l'autre	6
2.1.1.3	Coordination	8
2.1.2	Niveaux d'intention et théorie de contrôle de l'action	9
2.1.3	Application à la robotique interactive	11
2.2	Présentation de l'architecture actuelle	12
3	Le superviseur	15
3.1	Architecture du superviseur	15
3.2	Détails de l'architecture	17
3.2.1	Niveau Distal	17
3.2.1.1	Gestion des différents buts	17
3.2.1.2	Engagement des autres agents	17
3.2.1.3	Appel au planificateur	18
3.2.2	Niveau Proximal	19
3.2.2.1	Gestion de l'état du monde	19
3.2.2.2	Bon déroulement du plan	20
3.2.2.3	Gestion des signaux	20
3.2.3	Niveau Motor	23
3.2.3.1	Gestion des actions du robot	23
3.2.3.2	Monitoring de l'homme	24
3.2.3.3	Gestion de la tête du robot	25
3.3	Mise en œuvre	26
4	Perspectives et conclusion	28

Chapitre 1

Introduction

Ce rapport a pour but de faire le bilan du stage de fin d'étude que j'ai effectué au LAAS-CNRS au sein du thème robotique et plus particulièrement dans l'équipe Robotique et InteractionS (RIS).

La robotique est actuellement un domaine en pleine expansion et où de nombreuses avancées ont été faites ces dernières années. Cependant des progrès restent à faire et il reste aux robots actuels de multiples compétences à acquérir avant de devenir les robots compagnons du futur. La compréhension et la collaboration entre l'homme et le robot sont notamment des éléments clés qui demandent encore de nombreux développements. C'est sur cette thématique de la collaboration homme-robot que j'ai effectué mon stage.

L'objectif principal de mon stage a été de doter le robot de capacités permettant sa collaboration avec un homme. Pour faire cela il a fallu étudier et implémenter un ensemble minimal d'éléments permettant la compréhension et la collaboration entre l'homme et le robot. Le travail effectué durant le stage devait également permettre de développer une base sur laquelle il sera facile de travailler par la suite. Cette étude a été menée sur un exemple concret qui est la construction d'une pile d'objets en collaboration avec un homme.

1.1 Présentation du problème

L'exemple étudié a été tiré d'un workshop organisé lors de la conférence RO-MAN 2014 cet été. Ce workshop portait sur l'action jointe et a été organisé par des membres de mon équipe en collaboration avec des psychologues et des philosophes.

Lien vers le workshop : <http://fja2014.sciencesconf.org>

Le but de l'exemple est de construire une pile à l'aide de quatre cubes de couleurs différentes ainsi qu'un dernier cube blanc comme représenté figure 1.2. Au départ l'homme et le robot ont chacun à leur disposition une partie des cubes de couleur ainsi qu'un cube blanc (voir figure 1.1).

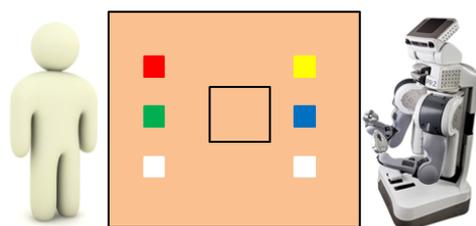


FIGURE 1.1 – État initial

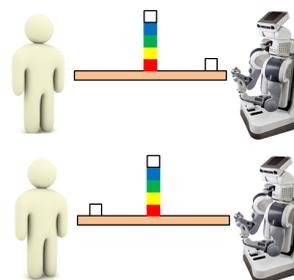


FIGURE 1.2 – Les deux états finaux possibles

Dans toute la suite de ce rapport l'homme et le robot seront appelés "agents". Afin de construire la pile les deux agents peuvent :

- Prendre un cube sur la table ou sur la pile.
- Poser un cube sur la pile ou sur la table.
- Donner un cube à l'autre agent.

Ces actions sont communes à l'homme et au robot, cependant il est important de garder à l'esprit qu'ils ne possèdent pas toujours les mêmes capacités.

Une fois les quatre cubes de couleur empilés arrive une phase de négociation pour savoir quel agent va poser le cube blanc.

Durant mon stage j'ai donc œuvré pour que le robot possède toutes les capacités utiles à la réalisation de ces actions, la compréhension de celles de l'homme et la coordination de ces dernières. Cependant, durant mon stage la phase de négociation a été supprimée, la pile finale n'est donc constituée que de quatre cubes de couleur.

Cet exemple a été implémenté sur le robot PR2 de Willow Garage et a été présenté lors du workshop.

Chapitre 2

Études préliminaires

J'ai dans un premier temps effectué un état de l'art sur l'action jointe et son application dans la robotique. Cet état de l'art a été fait non seulement dans le domaine de la robotique mais également dans ceux de la psychologie et de la philosophie, ces deux domaines étudiant eux aussi l'action jointe d'un point de vue collaboration homme-homme. En effet, l'étude de la collaboration entre hommes est un premier pas dans la réalisation d'une collaboration homme-robot car elle nous permet de comprendre les mécanismes nécessaires à cette collaboration.

2.1 État de l'art sur l'action jointe

Dans [SBK06] Sebanz et al. définissent l'action jointe de cette manière :

« Une action jointe est une forme d'interaction sociale où deux personnes ou plus coordonnent leurs actions dans le temps et l'espace pour apporter un changement dans l'environnement. »

Cette partie va donc détailler ce qui est nécessaire au bon déroulement de l'action jointe tel que le partage d'une intention, la prédiction des actions des autres agents ou la coordination. On retrouvera également dans cette partie le travail d'Elisabeth Pacherie sur le modèle de contrôle de l'action et les niveaux d'intention sur lesquels seront basés le travail de la partie suivante. Enfin on trouvera une partie dédiée à l'application de l'action jointe dans la robotique.

2.1.1 Théorie de l'action jointe, ce qui est nécessaire

Tomasello et al. définissent trois ingrédients nécessaires à une machine pour effectuer une action avec sa propre "intelligence" dans [TCC⁺05] :

- avoir un but vers lequel on veut agir.
- avoir la capacité d'agir pour modifier son environnement.
- avoir la capacité de percevoir son environnement afin de savoir quand le but est atteint.

Ils prennent l'exemple d'un thermostat :

- il a une température de référence à atteindre.
- il a la capacité d'éteindre ou d'allumer le chauffage afin de changer la température de la pièce.
- il possède un thermomètre lui permettant de connaître la température de la pièce.

Cependant, lorsque l'on effectue une action en collaboration avec un autre agent, ces trois capacités ne sont plus suffisantes. Plusieurs psychologues et philosophes ont tenté de lister ce qui est nécessaire à l'exécution d'une action jointe. Un des premiers à avoir abordé ce sujet est Bratman dans [Bra92]. Pour lui, les trois choses nécessaires à une action jointe sont :

- *L'attention mutuelle (mutual responsiveness)* : chaque agent doit essayer d'être réactif aux intentions et actions des autres agents tout en sachant que les autres le seront également. Chacun doit essayer d'agir en gardant un œil sur le comportement des autres, tout en sachant que les autres feront de même.
- *L'engagement dans l'action jointe (commitment to the joint activity)* : tous les agents sont engagés dans l'activité (pour des raisons qui peuvent être différentes) et la réponse mutuelle est dans la poursuite de cet engagement.
- *L'engagement dans un support mutuel (commitment to mutual support)* : chaque agent est prêt à aider les autres à effectuer leur rôle dans l'action jointe.

Plus tard Sebanz et al. proposent trois qualités nécessaires à une action jointe dans [SBK06] :

- *L'attention jointe* : savoir ce que les autres agents perçoivent (où ne perçoivent pas), partager une représentation de l'espace. Pouvoir diriger l'attention de l'autre et diriger notre propre attention vers ce que l'autre veut nous montrer.
- *Le partage des tâches* : prédire ce que les autres devraient faire et éventuellement pouvoir rentrer en négociation sur le partage des tâches. Ce partage se base sur la connaissance du rôle et des compétences des agents.
- *La coordination des actions* : prendre en compte les effets de nos actions et de celles des autres, se coordonner dans le temps et l'espace.

Enfin, Klein et al. dans [KWB⁺04] présentent 10 "challenges" nécessaires à un agent lors d'une action jointe :

1. Être engagé dans l'activité commune sous la forme d'une *convention basique* qui représente le niveau d'engagement de chaque participant pour soutenir la coordination.
2. Être capable de modéliser les autres agents de manière adéquate.
3. Être mutuellement prévisibles.
4. Pouvoir être dirigé.
5. Être capable d'exhiber, rendre visibles ses intentions à ses partenaires.
6. Être capable d'interpréter les signaux envoyés par ses partenaires.
7. Avoir les moyens de s'engager dans une négociation.
8. Ses méthodes de planification et son autonomie doivent permettre une approche collaborative du problème.
9. Être capable de participer à la gestion de l'attention des autres agents (porter à leur attention les signaux importants, etc.).
10. Pouvoir contrôler les coûts induit par les activités coordonnées.

Partant de là, trois éléments clés ont été identifiés et développés dans les parties suivantes :

- Le partage du but et de l'intention : chaque agent doit être engagé dans l'action et savoir que les autres le sont aussi.
- La perception et la prédiction des actions de l'autre : savoir ce que l'autre va faire et comment afin de plus facilement se coordonner.
- La coordination : cette coordination est à la fois spatiale et temporelle.

2.1.1.1 Le partage du but et de l'intention

L'un des premiers éléments nécessaire lors de la collaboration homme-robot est de partager un but et une intention concernant la tâche à effectuer, que chaque agent soit engagé dans l'action.

Tomasello et al. définissent dans [TCC⁺05] le **but** comme la représentation d'un état désiré et **l'intention** comme un plan d'action choisi et exécuté par l'agent dans la poursuite d'un but. Ce qui est appelé ici plan d'action est en fait l'ancrage de la tâche dans le temps, le fait que l'on ait prévu le moment d'agir. L'intention est donc liée à l'action que l'on veut effectuer (on a l'intention de faire A) tandis que le but est lié au résultat que l'on veut obtenir (on a pour but que l'état X soit satisfait). L'intention est un produit d'un but et d'un plan d'action construit en fonction de la perception de l'environnement et des connaissances/compétences de l'agent. Ce principe est illustré sur la figure 2.1.

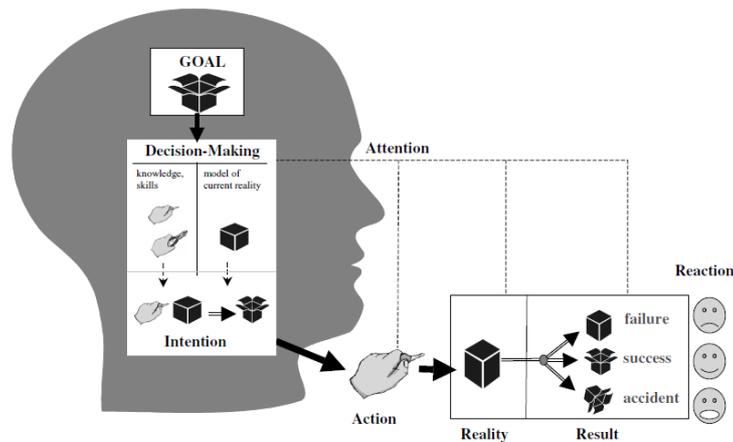


FIGURE 2.1 – Mécanisme de l'intention individuelle

De son côté Bratman dans [Bra93] définit le fait que nous avons l'intention de faire une action jointe J si :

1. J'ai l'intention que l'on fasse J et tu as l'intention que l'on fasse J.
2. J'ai l'intention que l'on fasse J en accord et à cause de (1) et des sous plans communs de (1). De même pour toi. C'est à dire que je vais choisir ma manière d'agir en te prenant en compte.
3. (1) et (2) sont des connaissances communes entre nous. C'est à dire que je sais que tu es engagé dans l'action et que ton plan d'action va me prendre en compte.

Tomasello et al. s'appuient ensuite sur Bratman pour représenter l'intention jointe figure 2.2. Pour eux, quand il s'agit d'une action jointe, le but doit alors contenir à la fois l'état désiré mais aussi le fait que l'action doit être effectuée avec l'autre agent. En ce qui concerne l'intention, les agents doivent choisir leur plan d'action en prenant compte le plan d'action de l'autre agent.

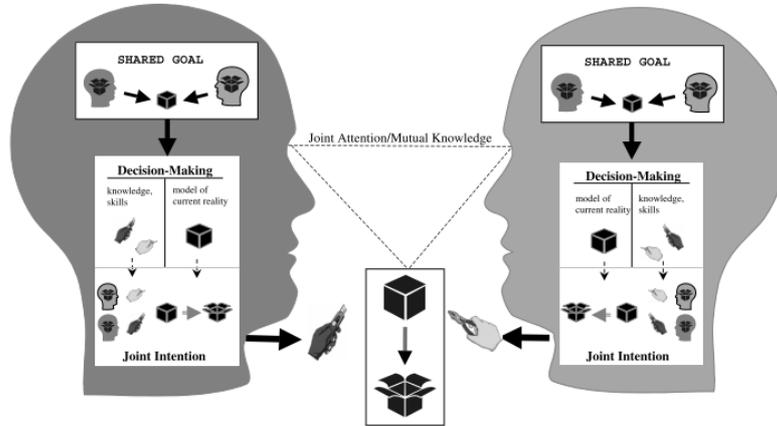


FIGURE 2.2 – Mécanisme de l'intention jointe

2.1.1.2 La perception et la prédiction des actions de l'autre

Dans une action jointe, il est important de savoir ce que les autres perçoivent ainsi que leurs capacités afin de pouvoir comprendre et prévoir leurs actions et leurs effets.

Dans [SBK06], Sebanz et al. détaillent trois outils essentiels permettant la perception lors d'une action jointe :

- *L'attention jointe* : La capacité à diriger son attention là où celle de son partenaire est dirigée permet d'avoir une représentation partagée des objets et événements. On saura alors ce que l'autre peut voir ou non, accéder ou non... Dans [Pac12] Pacherie rajoute qu'il doit y avoir manifestation mutuelle que l'attention est jointe, c'est à dire que l'on doit montrer à l'autre que nous partageons son attention.
- *L'observation de l'action* : Observer l'action de l'autre active une partie de notre cerveau appelée neurones miroirs qui va simuler la reproduction de cette action. En se basant sur nos propres connaissances de cette action et de la façon dont nous l'effectuons cela va donc nous permettre de prédire l'effet de cette action. Cela entre aussi dans un mécanisme appelé résonance moteur qui va favoriser la coordination, pas forcément consciente, lorsque deux agents effectuent la même action.
- *La co-représentation* : Avoir une représentation de l'autre, un modèle de son fonctionnement comprenant par exemple son but, ses capacités ou les règles sociales le dirigeant permet également de prédire ce qu'il peut/doit faire. Par exemple, un piéton qui voit un feu rouge va pouvoir prédire que les voitures vont s'arrêter.

Cette perception permet entre autre de prédire les actions des autres agents, cette prédiction peut être de deux types, décrites par Pacherie dans [Pac12] :

- *Prédiction "action-to-goal"* : cette prédiction utilise la résonance moteur définie précédemment, on va prédire l'objet d'une action que l'on observe. Par exemple si l'on observe quelqu'un qui tend le bras en direction d'un objet que l'on sait prenable, on

peut prédire qu'il va prendre cet objet. Le mot "goal" est ici considéré comme le but de l'action qui est observée.

- *Prédiction "goal-to-action"* : dans cette partie le mot "goal" représente maintenant le but global d'une tâche (par exemple construire une pile). Cette prédiction aura elle pour but de prédire la prochaine action de quelqu'un en connaissant son but et ses capacités.

Si l'on prend le cas de l'exemple traité pendant le stage ces prédictions peuvent être :

- *Prédiction "action-to-goal"* : On observe l'homme dirigeant son bras vers un cube, on en déduit qu'il s'apprête à prendre ce dernier.
- *Prédiction "goal-to-action"* : On sait que l'homme veut construire la pile avec le robot, on peut donc en déduire sa prochaine action telle que attraper un cube pour le poser sur la pile.

Enfin, dans [SK09], Sebanz et al. parlent de trois choses pouvant être déduites de ces prédictions et nécessaires à l'action jointe :

- *What* : de l'observation d'une action on peut comprendre ce que l'autre fait et prédire ce qu'il va faire après.
- *When* : le timing de l'action de l'autre permet d'aider à la coordination.
- *Where* : une action jointe s'effectue dans un espace partagé. Il est donc nécessaire de prédire les positions de son partenaire afin de pouvoir se coordonner dans l'espace et partager l'espace de travail.

2.1.1.3 Coordination

Dans [KBS11], Knoblich et al. définissent deux sortes de coordination :

- **La coordination émergente** : Elle n'est pas volontaire et est indépendante de tout plan commun ou connaissance commune. Elle est due à la perception couplée d'une action et va amener des individus à agir de la même façon. Il y a quatre processus qui peuvent amener à une coordination émergente :
 - *L'entraînement* : C'est un processus qui va amener la coordination temporelle de deux acteurs en l'absence de liens mécaniques. Par exemple, des études ont montrées que des personnes assises sur des rocking chairs différents finissent par se balancer à la même fréquence.
 - *Les affordances communes et jointes* : Les affordances (voir [Gib79]) sont les possibilités que nous donne un objet pour une certaine action (par exemple les différents moyens de tenir une tasse). Une affordance commune représente le cas où plusieurs agents ont le même répertoire d'action et perçoivent le même objet, il vont donc agir de la même manière. Une affordance jointe représente le cas où un objet ne possède d'affordance que pour deux personnes, si l'on veut effectuer l'action il y aura alors nécessairement coordination (par exemple une scie passe partout, scie avec une poignée à chaque extrémité).
 - *La correspondance action-perception* : Comme dit dans la partie précédente, l'observation d'une action peut amener à effectuer la même action de manière non volontaire. Sur ce principe, il peut y avoir coordination si plusieurs personnes observent la même action qui va faire apparaître chez eux la même réaction.
 - *Simulation de l'action* : Comme expliqué précédemment l'observation d'une action amène à la prédiction des effets de cette action. Si plusieurs personnes observent la même action et prédisent le même effet ils peuvent donc être amenés à avoir la même réaction en retour (par exemple quand deux personnes observent un même objet tomber elles vont toutes les deux avoir la réaction de chercher à le rattraper).
- **La coordination planifiée** : C'est le cas quand les différents agents prennent en compte les autres pour planifier leur action, la coordination est volontaire. Elle nécessite une représentation partagée de la tâche et une perception jointe. Pour aider à cette coordination il peut être nécessaire de la part des agents de changer leur comportement par rapport à une action qu'ils effectueraient seuls. Ces changements de comportement sont ce que Vesper et al. appellent *coordination smoothers* dans [VBKS10]. Ils en donnent plusieurs types possibles :
 - Rendre son comportement plus prévisible : faire des gestes plus amples ou alors moins variables...
 - Délimiter et structurer sa propre tâche pour réduire le besoin de coordination : partager l'espace de travail, travailler à tour de rôle...
 - Produire des signaux de coordination : regarder quelqu'un quand c'est son tour d'agir, compte à rebours...
 - Se synchroniser : agir au même rythme que l'autre, agir de façon régulière...
 - Changer l'utilisation des objets : un objet peut avoir différentes affordances suivant si il est utilisé seul ou en collaboration.

2.1.2 Niveaux d'intention et théorie de contrôle de l'action

Cette partie va présenter le travail effectué par Elisabeth Pacherie sur la modélisation de l'action. Ce travail servira de base au développement de l'architecture dans la suite du stage.

Dans [Pac12] Pacherie explique que pour elle une action peut être divisée en trois niveaux d'intention ayant chacun un rôle distinct dans le monitoring de l'action. Ces trois niveaux sont les suivants :

- **Distal Intention** : C'est le niveau le plus haut, à ce niveau on raisonnera en fonction du but et de l'environnement global de l'action. Il décidera des moyens d'effectuer l'action (création d'un plan d'action).
- **Proximal Intention** : Ce niveau hérite du plan d'action créé par le précédent. Il a pour but de superviser l'exécution de ces tâches dans le temps. Il décidera quand commencer une tâche et surveillera son bon déroulement.
- **Motor Intention** : Ce niveau a comme rôle la gestion d'une tâche qu'il reçoit du niveau précédent. Il aura une perception du monde relative à la tâche et effectuera les choix locaux relatifs à cette tâche.

Elle représente la coordination entre ces trois niveaux sur la figure 2.3, ces niveaux sont agencés en cascade.

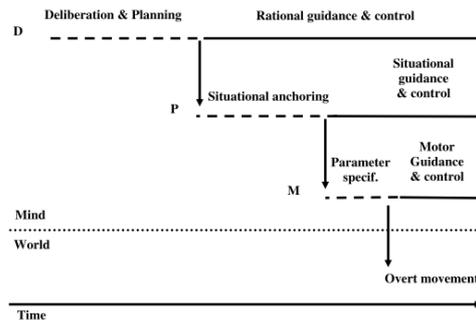


FIGURE 2.3 – Coordination des trois d'actions

Dans [Pac12] elle reprend ces trois niveaux d'intention et les étend au cas d'une action jointe. Ils deviennent alors :

- **Shared Distal Intention** : Le plan pris en compte à ce niveau n'est plus forcément le plan global mais un sous plan ainsi que les parties communes des sous plans des autres. Ce qui est représenté à ce niveau pourra être effectué par d'autres agents.
- **Shared Proximal Intention** : Ce niveau requiert six choses :
 - **Self-prediction** : Chaque agent représente ses actions et ses conséquences.
 - **Other-prediction** : Chaque agent représente les actions des autres et leurs conséquences.
 - **Dyadic ajustement** : Chaque agent représente comment ils vont affecter les autres et vice-versa et ajuste leurs actions en fonction.
 - **Joint action plan** : Chaque agent a une représentation de la hiérarchie des buts et des états résultants.

- **Joint predictions** : Chaque agent prédit les effets joints de leur propre action et de celles des autres.
- **Triadic adjustment** : Chaque agent utilise la perception jointe pour gérer leur progrès et décider de leur prochain mouvement.

Pour cela ce niveau a besoin de trois choses définies précédemment : l'attention jointe, la co-représentation et la prédiction.

- **Coupled Motor Intention** : Il est nécessaire que les actions soit coordonnées précisément dans le temps et l'espace. Ce niveau devra donc le prendre en compte.

Toujours dans [Pac08] Pacherie reprend un modèle théorique de contrôle de l'action. Ce modèle est présenté sur la figure 2.4.

Dans ce modèle on retrouve deux modèles relatifs aux deux types de prédiction précédemment définis :

- **Inverse models** : Retourne les commandes nécessaires pour obtenir un état (Goal-to-action prediction).
- **Forward models** : Retourne une prédiction des conséquences d'une action (Action-to-goal prediction).

Ce modèle utilise trois comparateurs :

- entre l'état désiré et l'état prédit afin de maintenir des performances précises en présence de délais de retour d'information.
- entre l'état prédit et l'état estimé actuel pour mettre à jour le modèle direct (forward models) et améliorer son fonctionnement.
- entre l'état désiré et l'état actuel estimé pour mettre à jour le modèle inverse et améliorer son fonctionnement.

Elle associe ensuite ce modèle à chacun des trois niveaux d'intention définis précédemment et les obtient le modèle figure 2.5

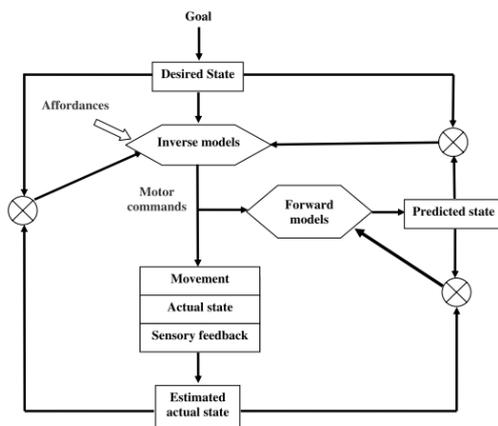


FIGURE 2.4 – Modèle théorique de contrôle de l'action

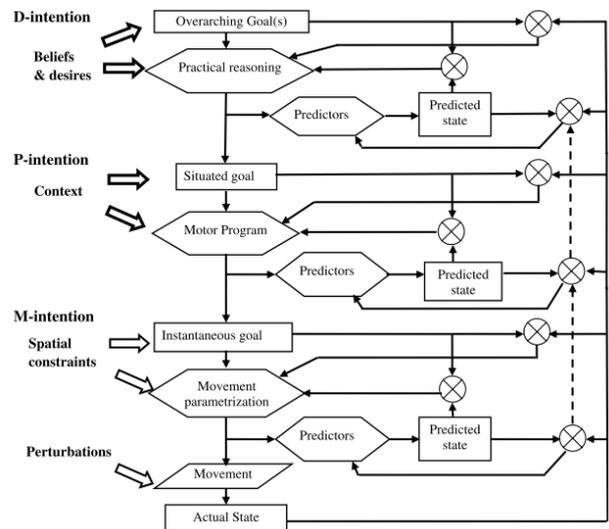


FIGURE 2.5 – Modèle 3 niveaux du contrôle de l'action

2.1.3 Application à la robotique interactive

Plusieurs groupes de roboticiens ont effectué des travaux portants sur différents aspects de l'action jointe.

Par exemple dans [GBB⁺05] Gray et al. présentent un travail basé sur les neurones miroirs effectué sur le robot Leonardo. Leur idée est de doter Leonardo d'un système inspiré des neurones miroirs lui permettant de comprendre le but d'une action qu'il observe. Pour faire cela il va convertir les mouvements qu'il observe dans une base relative à ses propres mouvements afin des les reconnaître. Ensuite, en se basant sur la perception que le robot sait que l'homme possède et du but de la tâche il va chercher les actions possible correspondant à la suite de mouvement qu'il observe. Une fois l'action trouvée il pourra alors prédire le comportement de l'homme et possiblement réagir.

Toujours concernant le prédiction des actions de l'homme Hoffman présente dans [Hof10] un procédé construit autour d'une extension des chaînes de Markov permettant d'obtenir la probabilité concernant la prochaine action. Ce modèle est amélioré de sorte à ce qu'au fur et à mesure du temps les probabilités évoluent en fonction des suites d'actions déjà effectuées (la probabilité d'une suite d'action sera plus forte si elle a déjà été effectuée).

Concernant le communication entre l'homme et le robot, Hoffman et Breazeal ont implémentés dans [HB04] un set de moyens de communications visuelle sur le robot Leonardo permettant de faire comprendre son état d'esprit. Dans [Bre02], Breazeal travail sur les signaux échangés pour gérer le *turn-taking* dans une conversation homme-robot.

Peu de travaux sur l'architecture robotique prennent l'homme en compte. Dans [FSS⁺06] Fong et al. présentent une architecture nommée HRI/OS permettant aux robots et aux hommes de travailler en équipe. Le système permet de dispatcher les tâches en prenant en compte les différentes capacités des agents. La communication entre les agents n'est pas directe, quand un agent a besoin d'aide pour effectuer un tâche il demande au système qui va lui fournir l'agent le plus apte à l'aider disponible.

2.2 Présentation de l'architecture actuelle

Afin d'interagir avec le monde extérieur ainsi qu'avec l'homme le robot a besoin de nombreuses capacités telle que la planification de mouvements, la compréhension de son environnement, etc... Une architecture permettant de faire cela a été développée au LAAS ([ACF⁺98]). Cette architecture est composée de trois niveaux avec lesquels on peut faire l'analogie avec les trois niveaux d'intention de Pacherie :

- **Le niveau de décision :** Ce niveau permet de produire un plan et de le superviser. Il est donc composé d'un planificateur ainsi que d'un superviseur. Il envoie au niveau d'exécution les différentes actions à effectuer et analyse les rapports d'exécution qu'il lui retourne. Il correspond au niveau *Distal Intention* de Pacherie.
- **Le niveau d'exécution :** Il permet de contrôler et exécuter les différentes actions. Il reçoit du niveau de décision les actions à effectuer et lui retourne des rapports d'exécution. Il est composé de deux parties : une première gérant les requêtes afin de les paramétrer et de les transférer aux modules du fonctionnel et une seconde gérant les flots de données reçus du fonctionnel afin d'en faire des rapports pour le niveau de décision. Il correspond au niveau *Proximal Intention* de Pacherie.
- **Le niveau fonctionnel :** Il contient les éléments fonctionnels et de perception du robot. Il est découpé en modules implémentés en Genom (voir [FHC97]). Un module gère une ressource. Il reçoit en entrée des paramètres et ordres du niveau d'exécution. En retour, il renvoie un rapport d'exécution et possiblement des données. Chaque module est composé de deux parties : une première permettant de gérer le module par rapport à son état et aux requêtes qu'il reçoit, et la seconde permettant d'effectuer l'activité demandée. Ce niveau correspond au niveau *Motor Intention* de Pacherie.

Sur cette base, l'architecture implémentée sur le robot est la suivante :

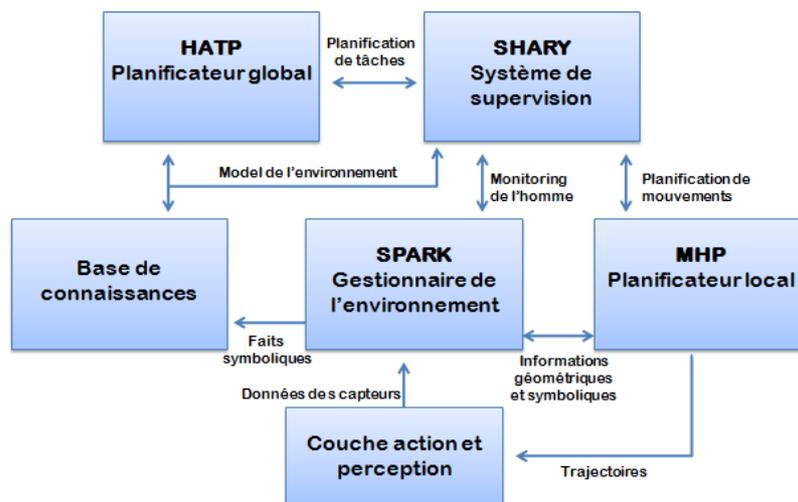


FIGURE 2.6 – Architecture globale

Les composants de cette architecture sont les suivants :

- **Le superviseur, SHARY (Supervision for Human Adapted Robot Y(I)nteraction)** : c'est le plus haut niveau de l'architecture, c'est lui qui est en charge du bon déroulement des actions du robot et de son interaction avec l'extérieur. Il détermine quels sont les buts à atteindre et il fait appel au planificateur pour obtenir un plan relatif à ces buts dont il veillera par la suite au bon déroulement à l'aide des rapports des autres modules. Il permet également le monitoring des actions de l'homme. Il doit être assez robuste pour pouvoir s'adapter aux différents échecs ou aux comportements inattendus de la part de l'homme. C'est sur ce superviseur que j'ai été amenée à travailler durant mon stage (voir [FCA14] et [CAM09]).
- **Un planificateur global, HATP (Hierarchical Agent-Based Task Planner)** : il permet de produire un plan, relatif à un but donné (voir [AWG⁺11]). La planification permet de prendre en compte non seulement le robot mais aussi d'autres agents participant à la tâche, le plan retourné contient donc à la fois des actions à effectuer par le robot, d'autres pour les autres agents mais aussi certaines communes à plusieurs agents appelées actions jointes. Un coût et une durée sont associés à chaque action, mais HATP peut également planifier en fonction de certaines contraintes sociales telles que :
 - éviter des états indésirables à l'homme.
 - éliminer les séquences d'action qui pourraient être mal interprétées par l'homme.
 - ajuster les efforts entre les différents agents.
 - éviter des grands temps d'attente entre deux actions de l'homme.
 - limiter les dépendances entre les actions de deux agents ou plus.
- **Un gestionnaire des connaissances sur l'environnement, SPARK (Spatial Reasoning and Knowledge)** : Le fonctionnement de SPARK est décrit dans [AWG⁺11], [WGLA12] et [MWCA14]. Il récupère les informations sur son environnement grâce à différents capteurs. En ce qui concerne les hommes, on récupère leurs positions et postures à l'aide de kinects (cameras permettant la reconnaissance 3D). Pour le robot, ces mêmes données sont récupérées en interne à l'aide de ROS (OS de robotique). Enfin, les différents objets de la scène sont repérés par des tags à l'aide des caméras présentes sur le robot.

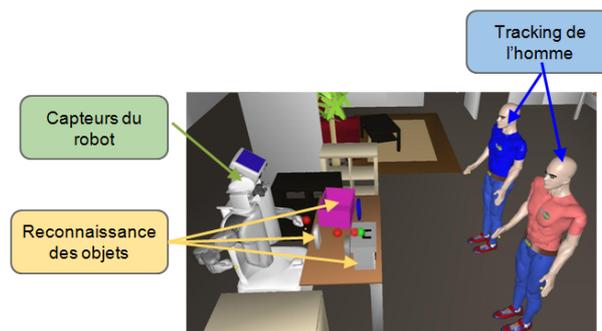


FIGURE 2.7 – Récupération des données par SPARK

Une fois toutes les données récupérées, SPARK calcule des faits symboliques entre les objets et entre les objets et les agents. Ces faits sont par exemple :

- des positions relatives (cassette sur la table, dans une boîte, ...).
- des capacité de perception ou de manipulation des agents (cassette visible par robot, cassette accessible par l'homme, ...).

Les faits ainsi calculés sont stockées dans la base de connaissance.

SPARK permet également de monitorer les actions de l'homme en analysant à tous moments la position de ses mains dans l'environnement.

SPARK gère également les croyances des différents agents. En effet, un agent peut faire des hypothèses sur la position d'un objet et donc avoir une croyance quand à cette position différente de la réalité. Pour faire cela SPARK calcule les faits décrits précédemment en se mettant à la place de tous les agents (perspective taking). On obtient donc un modèle de l'environnement du point de vu de chaque agent présent dans la scène.

- **Une base de connaissance** : c'est dans cette dernière que SPARK stocke les faits qu'il calcule. Elle permet au superviseur de pouvoir à tout moment connaître l'état du monde mais aussi à HATP de planifier en se basant sur ces faits.
- **Un planificateur de mouvements, MHP (Manipulation in Human Presence)** qui permet de calculer les trajectoires du robot pour une action donnée en fonction de son environnement et des autres agents (voir [SLL01]).
- **Une couche action et perception** qui permet la gestion des capteurs et actionneurs du robot.

Chapitre 3

Le superviseur

3.1 Architecture du superviseur

Un superviseur fonctionnel était déjà présent à mon arrivée en stage. Mon travail a consisté à le structurer de manière à construire une architecture fonctionnelle, correspondant aux modèles théoriques et facilement extensible ainsi que de rajouter des éléments nécessaires à la collaboration homme-robot non présents initialement.

De même que l'architecture générale du système, l'architecture du superviseur développée est basée sur les trois mêmes niveaux d'intention de Pacherie. Si l'on fait correspondre les trois niveaux d'intention proposés par Pacherie dans le cas homme-homme au cas homme-robot on peut décrire les trois niveaux de la manière suivante :

- **Distal** : A ce niveau, comme le propose Pacherie, le robot raisonne en fonction des buts et de l'environnement global de l'action. C'est à ce niveau que le robot décide quel but il effectue et à quel moment. C'est également ici que l'on estime le niveau d'engagement des autres agents pour ces différents buts. Enfin, le niveau distal fait appel à HATP afin d'obtenir un plan correspondant au but qu'il a choisi. Il transmet ensuite ce plan au niveau proximal.
- **Proximal** : Ce niveau reçoit un plan envoyé par le niveau supérieur. Il est en charge de veiller au bon déroulement de ce plan. Pour faire cela c'est lui qui a la charge d'affiner ces actions dans le contexte et décider à quel moment effectuer les actions comprises dans le plan qu'il envoie ensuite au niveau motor. Afin d'effectuer certaines de ces actions, il ce niveau aura aussi la charge de gérer l'échange des signaux entre l'homme et le robot spécifiques à l'action. Il veille ensuite au bon déroulement de ces actions en analysant les rapports reçus du niveau motor. A ce niveau on maintient un état du monde relatif à la tâche en cours.
- **Motor** : Ce niveau à la charge de la bonne exécution des actions du robot et du monitoring de celles de l'homme. Il reçoit ces actions du niveau proximal, fait appel aux différents modules permettant leurs exécutions ou leurs monitorings dans le cas de l'homme et renvoie des rapports concernant leurs déroulements.

On obtient donc l'architecture suivante :

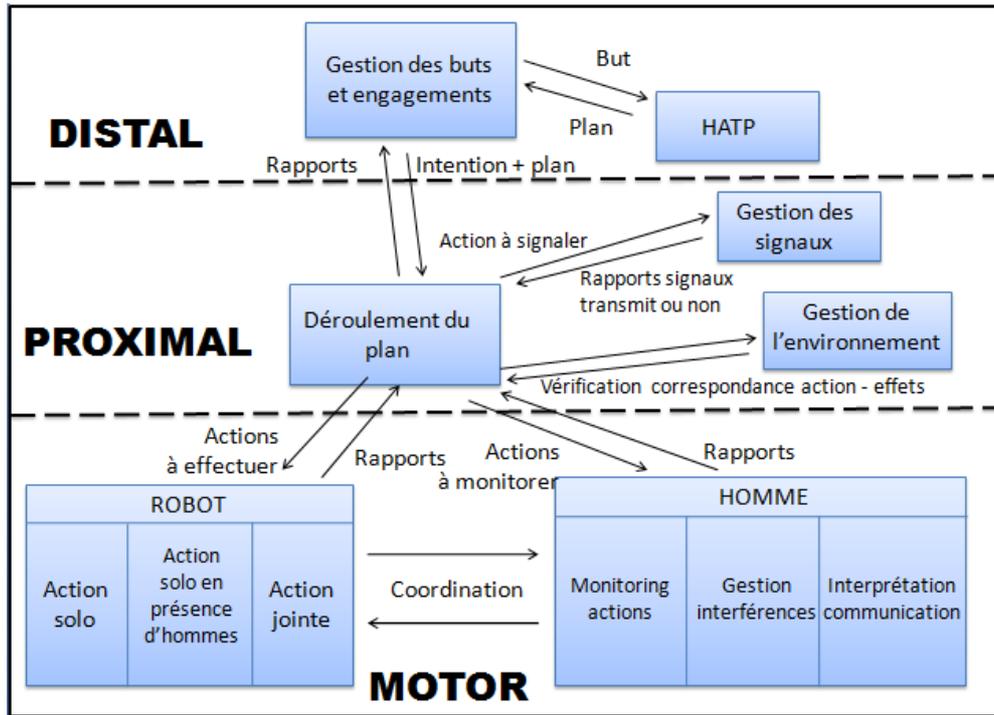


FIGURE 3.1 – Architecture superviseur

Le détail de cette architecture et de ce qui la compose est décrit dans la partie suivante.

Dans cette partie, je vais expliquer la déclinaison que j'ai effectué de ces trois niveaux d'intention, en montrant les éléments importants, les questions posées, les réponses que j'y ai apporté et les pistes d'étude ouvertes.

3.2 Détails de l'architecture

3.2.1 Niveau Distal

Dans ce niveau on va retrouver trois éléments : la gestion des différents buts arrivant au robot, la gestion de l'engagement des différents agents et le lien avec le planificateur HATP.

3.2.1.1 Gestion des différents buts

- *Problématique* : Même si dans l'exemple traité la tâche à effectuer est unique et définie précisément ce ne sera pas toujours le cas dans toutes les applications possibles de la robotique. En effet, un robot autonome doit être capable de décider quelle tâche effectuer et quand. Il doit également être capable de stopper une tâche en cours si un ordre plus important arrive.
- *Solution implémentée* : Actuellement les seules arrivées de buts possibles pour le superviseur sont soit une entrée manuelle de la tâche à effectuer, soit des ordres donnés par l'utilisateur à l'aide d'une tablette. Une gestion de ces deux types d'entrées était déjà présente dans le superviseur et a donc été intégrée dans l'architecture. Dans cette solution la priorité va toujours aux ordres reçus par la tablette. En ce qui concerne les ordres rentrés manuellement, le robot effectuera les ordres dans l'ordre de réception et jusqu'à leur fin ou leur échec. Si entre temps un ordre arrive de la tablette le robot va, soit abandonner la tâche en cours si l'ordre est une action différente ou celui d'abandonner la tâche, soit suspendre la tâche en cours si l'ordre est de se mettre en pause.
- *Améliorations possibles* : Une amélioration possible à apporter à cette solution est la gestion de la reprise d'une tâche qui aurait été interrompue par un ordre de la tablette. Pour faire cela il faudrait pouvoir déterminer si après l'exécution de l'ordre venant de la tablette la tâche précédente est encore faisable et s'il est toujours nécessaire de la réaliser. Une autre amélioration à apporter serait la reconnaissance d'une tâche effectuée par l'homme. Il serait bien de pouvoir, grâce à l'observation des actions de l'homme, déduire quelle tâche il est entrain d'effectuer et s'il est possible de lui venir en aide. Enfin, il pourrait être bien de posséder un "agenda" contenant les différents buts que le robot a à effectuer, il permettrait de décider quel but est prioritaire et quand commencer chaque tâche.

3.2.1.2 Engagement des autres agents

- *Problématique* : Quand on effectue une tâche en collaboration avec un autre agent il est nécessaire de savoir si l'autre agent est réellement impliqué dans la tâche ou non. En effet, quand un robot travaille avec un homme il se peut que l'homme choisisse d'abandonner la tâche en cour de route où alors que le robot ait mal interprété ce que voulait faire l'homme et donc qu'il n'essaye pas d'effectuer la même tâche que lui. Il est donc nécessaire de savoir détecter si l'homme est intéressé ou non (attentif, actif...) et également s'il participe de manière utile à la tâche où s'il agit différemment.
- *Solution implémentée* : La perception de l'homme mise en œuvre actuellement ne permettant pas de détecter si l'homme est attentif ou non (orientation de la tête principalement) le travail implémenté pour cette partie a donc été centré sur l'interprétation des actions de l'homme. En effet, au commencement d'une tâche le robot calcule un plan contenant des actions

pour l'homme et pour lui même, cependant il n'y a aucune certitude sur le fait que l'homme va effectivement effectuer les actions prévues et au moment souhaité. On pourra donc recevoir l'information que le plan a échoué car l'homme n'a pas agi de la manière requise. Cependant cela ne voudra pas forcément dire que l'homme n'est pas impliqué dans la tâche, il peut par exemple vouloir l'effectuer d'une manière différente de ce qui a été prévu. Il est donc nécessaire de déterminer pour chaque action si elle est dans le plan ou non mais aussi pour les action non contenues dans le plan si elle est "mauvaise" ou non (ne fait pas avancer la tâche, la fait reculer).

La solution choisie pour faire cela est la suivante : lorsque l'on reçoit l'information que l'homme a effectué une action non contenue dans le plan on redemande alors un nouveau plan au planificateur et on compare le nombre d'actions (pondéré d'un certain coût) contenues dans ce plan par rapport à celles restantes dans le plan initial. Si ce nombre est plus élevé (la tâche est alors considérée comme plus difficile à effectuer après l'action de l'homme) ou si le planificateur n'a pas pu trouver de nouveau plan on considère alors l'action de l'homme comme "mauvaise". On pourra alors ensuite fixer un nombre maximal de "mauvaises" actions à partir duquel on considérera que l'homme n'est plus engagé dans la tâche.

- *Exemple* : Dans l'exemple traité une "mauvaise" action est par exemple de poser le mauvais cube sur la pile ou alors de retirer un cube correct de la pile. Par contre, si par exemple un cube est accessible à la fois par le robot et par l'homme et que dans le plan du robot c'est à lui de le poser sur la pile si l'homme le fait à sa place c'est alors une action non contenue dans le plan mais qui n'est pas "mauvaise".
- *Améliorations possibles* : Cette manière de déterminer si une action est "mauvaise" ou non est une première étape mais n'est pas suffisante en soi. En effet, pour une tâche compliquée l'homme ne pourra pas forcément trouver le plan optimal aussi facilement que le robot. Il pourra donc être amené à effectuer des actions qu'il pense bonnes mais qui en réalité rallongent l'exécution du plan calculé par le robot. Il est donc nécessaire de prendre d'autres critères en compte lors de cette détermination. Le nombre maximal de "mauvaises" actions pourrait également être remplacé par la gestion d'un degré d'engagement qui varierait en fonction des "bonnes" et "mauvaises" actions de l'homme ainsi qu'en fonction des signaux envoyés par ce dernier.

3.2.1.3 Appel au planificateur

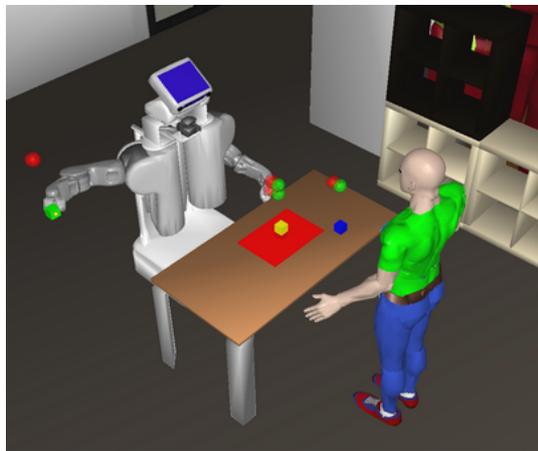
- *Le lien entre le superviseur et HATP* : Il est nécessaire d'avoir à l'intérieur du superviseur une interface avec le planificateur. Cette interface permet à la fois de fournir au planificateur les données nécessaires à la planification contenues dans la base de connaissance mais aussi de gérer les requêtes faites au planificateur et ses réponses. Cette interface était déjà présente dans le superviseur initial et a été réutilisée.
- *Le domaine HATP spécifique à l'exemple* : Pour calculer un plan, HATP nécessite que l'on crée un domaine correspondant à la tâche effectuée. Ce domaine sera principalement la description des actions possibles pour réaliser cette tâche associées à leurs pré-conditions (ce qui est nécessairement dans l'état du monde avant cette action) et leurs effets (ce que l'action implique sur l'état du monde). On écrira également dans ce domaine sur quels faits de la base de connaissance le planificateur va se baser (transmit à l'aide du superviseur) et quel est l'état final du monde voulu. Il a donc été nécessaire de créer un nouveau domaine associé à l'exemple traité. Ce domaine permet la planification depuis n'importe quel état initial (pile initialement fautive, un agent déjà en possession d'un des cubes...).

3.2.2 Niveau Proximal

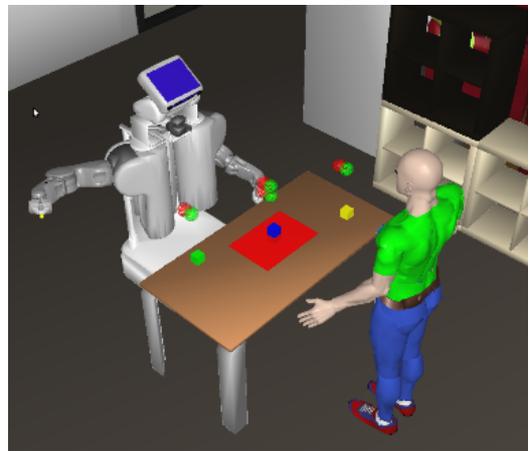
Dans ce niveau on va retrouver trois éléments également : la gestion de l'état du monde relatif à la tâche, la gestion du bon déroulement du plan ainsi que la gestion des signaux échangés avec l'homme.

3.2.2.1 Gestion de l'état du monde

- *Problématique* : Lors de l'exécution d'une tâche il est nécessaire d'avoir une connaissance de l'environnement relative à cette tâche, connaître par exemple les positions des objets concernés par cette tâche ou bien l'avancement de cette dernière. Pour l'exemple traité il est utile de connaître où en est la construction de la pile, si les cubes la composant sont dans un ordre correct...
- *Solution implémentée* : Une solution spécifique à l'exemple a été développée. Après chaque action de l'homme ou du robot le superviseur va déterminer pour chaque cube sa position grâce aux faits retournés par SPARK (cube sur la table, cube rouge sur le cube vert...). Cette position peut être de quatre types : le cube est sur la table, le cube est dans la pile, le cube est tenu par le robot ou alors le cube est tenu par l'homme. En ce qui concerne la pile, il déduira des mêmes faits de quels cubes elle est constituée et dans quel ordre ils sont disposés. Il comparera ensuite cet ordre avec l'ordre prédéfini afin de savoir si la pile est correcte ou non. C'est ces informations qui serviront lors des vérifications des post-conditions suivant une action (voir paragraphe suivant). Par exemple après l'ajout d'un cube on va vérifier que ce cube est bien dans la pile et à son sommet, ou alors après la prise d'un cube par un agent que le cube est effectivement en sa possession.
- *Exemple* :



state red-cube pile
state yellow-cube pile
state green-cube robot
state blue-cube table
order pile red-cube yellow-cube
state pile ok
top pile green-cube



state red-cube pile
state yellow-cube table
state green-cube table
state blue-cube pile
order pile red-cube blue-cube
state pile wrong
top pile blue-cube

- *Améliorations possibles* : Cette partie du superviseur nécessite d'être généralisée à toutes les tâches possibles. Une manière de faire cela serait par exemple de faire un lien direct avec le domaine HATP afin de savoir pour chaque action quelles sont les pré-conditions nécessaires et quels effets elle doit avoir.

3.2.2.2 Bon déroulement du plan

- *Problématique* : Le niveau proximal reçoit un plan du niveau distal, il est important de veiller à son bon déroulement. Pour faire cela il est nécessaire dans un premier temps de dispatcher les différentes actions du plan entre l'homme et le robot. Pour le robot, il est nécessaire pour chaque action de déterminer à quel moment l'exécuter, c'est à dire quand est ce que les pré-conditions associées à cette action sont vérifiées. Pour l'homme on doit s'assurer qu'il effectue bien les actions prévues aux moments adéquats. Il est également nécessaire pour chaque action de vérifier que cette action se déroule correctement et que l'état du monde après l'action est bien celui attendu.
- *Solution implémentée* : Cette partie était déjà présente dans le superviseur est a donc été intégrée à l'architecture. La vérification des post-conditions des actions à l'aide de la gestion de l'état du monde décrit ci-dessus a été ajoutée.
- *Améliorations possibles* : Actuellement si quelque chose ne se passe pas comme prévu lors du déroulement du plan (échec, action inattendue de l'homme...) le plan échoue et une re planification est alors nécessaire. Il serait bien de pouvoir soit effectuer une action de rattrapage, soit pouvoir replanifier seulement pour une partie plus petite de la tâche.

3.2.2.3 Gestion des signaux

- *Problématique* : Lorsque l'on effectue une action en collaboration avec un homme il est nécessaire de lui transmettre des signaux et de comprendre ceux qu'il nous renvoie. Pour la transmission des signaux deux questions se posent alors : quand les transmettre et comment ?
- *Début de solution* : Cette partie du superviseur n'est pas encore implémentée, cependant un début de réflexion a été effectué.
Dans un premier temps il est utile de connaître un certain nombre de faits relatifs aux actions. Ces faits sont de deux sortes :
 - Une partie concernant les actions effectuées :
 - * *agent hasDone action parameters* : sera stocké pour l'agent qui aura effectué l'action.
 - * *agent hasSeen action parameters* : sera stocké pour tous les agents présents et attentifs lors de l'exécution de l'action.
 - * *agent hasDeduct action parameters* : sera stocké pour un agent qui n'était pas présent au moment de l'action mais qui a pu observer l'effet de cette action sur l'environnement.
 - * *agent ignore action parameters* : sera stocké pour les agents qui n'auront ni vu l'action ni son effet sur l'environnement.

On pourra ainsi déduire les faits suivants :

- * *agent knowAuthor author action parameters* : si l'agent a fait ou vu l'action.
- * *agent thinkAuthor author action parameters* : si l'agent a déduit l'action et a un moyen de présumer de l'auteur de cette action (un seul agent susceptible de l'avoir effectuée).
- * *agent ignoreAuthor author action parameters* : si l'agent ignore l'action ou l'a déduit et n'a aucun moyen de savoir qui l'a effectuée.
- * *agent knowDetails action parameters* : si l'agent a fait ou a vu l'action. L'agent connaît alors la manière dont a été faite l'action.

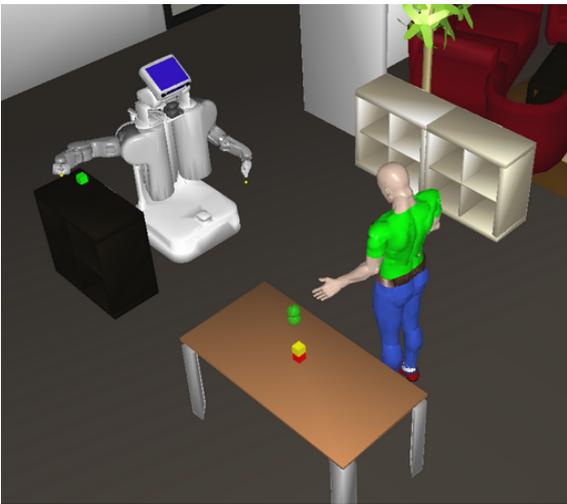
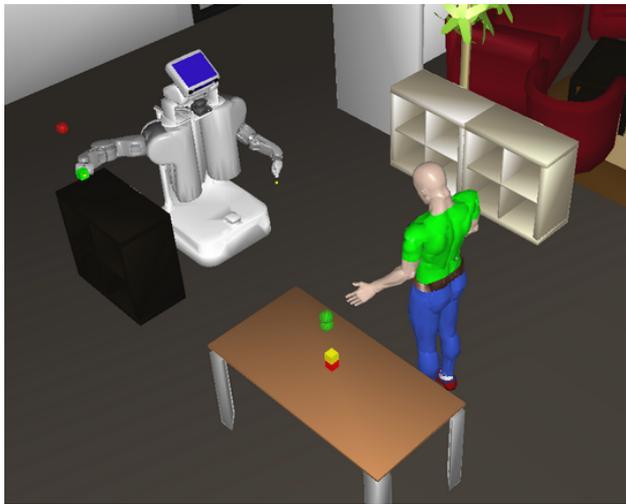
Voici une illustration de ces faits sur un exemple simple :

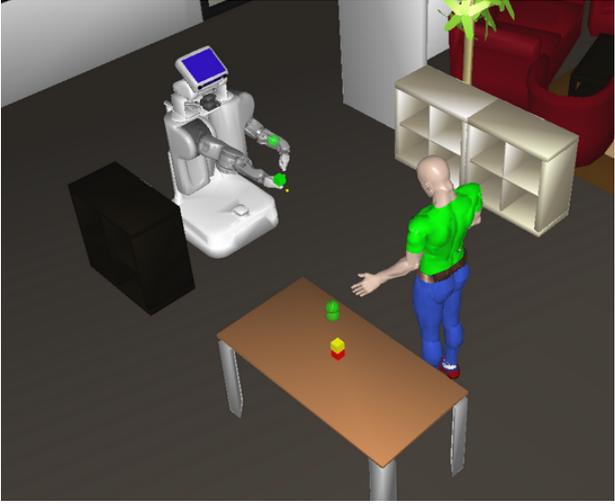
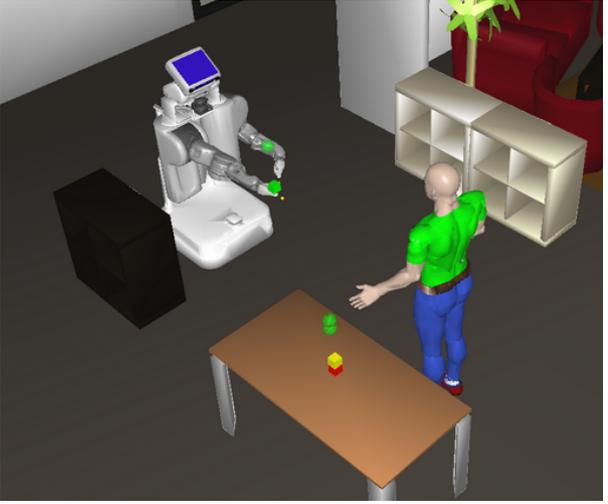
<p>PR2 pose le cube jaune sur le cube rouge, GREEN-HUMAN est attentif et observe l'action, BLUE-HUMAN ne peut pas voir l'action.</p>	<p>BLUE-HUMAN observe maintenant la scène, il peut observer que quelqu'un a mis le cube rouge sur le jaune mais ne peut pas deviner qui.</p>
	
<p>PR2 hasDone place yellow-cube red-cube GREEN-HUMAN hasSeen place yellow-cube red-cube BLUE-HUMAN ignore place yellow-cube red-cube PR2 knowAuthor PR2 place yellow-cube red-cube GREEN-HUMAN knowAuthor PR2 place yellow-cube red-cube BLUE-HUMAN ignoreAuthor place yellow-cube red-cube PR2 knowDetails place yellow-cube red-cube GREEN-HUMAN knowDetails place yellow-cube red-cube</p>	<p>BLUE-HUMAN hasDeduct place yellow-cube red-cube</p>

– Une partie concernant les signaux échangés et le déroulement de l'action, utile lors d'une action jointe :

- * *agent wantTo action parameters* : un agent veut effectuer l'action.
- * *agent isReadyTo action parameters* : un agent est prêt à effectuer l'action.
- * *agent1 sentSignal agent2 action parameters* : l'agent 1 a envoyé un signal à l'agent 2 un signal pour effectuer l'action.
- * *agent1 receiveSignal agent2 action parameters* : l'agent 1 a reçu un signal de la part de l'agent 2 pour effectuer l'action.

Voici une application de ces faits sur l'exemple d'un robot voulant donner un cube à un homme :

<p>PR2 veut donner le cube vert à GREEN-HUMAN pour qu'il le pose sur la pile.</p>	<p>PR2 a maintenant le cube en main, il est prêt à le donner.</p>
	
<p>PR2 wantTo give GREEN-HUMAN green-cube</p>	<p>PR2 isReadyTo give GREEN-HUMAN green-cube</p>

<p>PR2 signifie à GREEN-HUMAN qu'il est prêt GREEN-HUMAN est attentif, il reçoit donc le signal.</p>  <p>PR2 sentSignal GREEN-HUMAN give green-cube GREEN-HUMAN receiveSignal PR2 give green-cube</p>	<p>GREEN-HUMAN montre son accord il est donc prêt, l'action peut commencer.</p>  <p>GREEN-HUMAN sentSignal grab green-cube GREEN-HUMAN wantTo grab PR2 green-cube GREEN-HUMAN isReadyTo grab PR2 green-cube</p>
--	---

Il faut ensuite décider *quand* et *comment* transmettre les signaux à l'homme.

En ce qui concerne le *quand* il y a deux cas possibles :

- L'homme est attentif : on peut alors transmettre le signal dès que l'on est prêt à effectuer l'action.
- L'homme n'est pas attentif : on attend alors qu'il prête son attention au robot. Si la tâche à effectuer est de caractère urgent on peut interpellé l'homme (appel vocal).

Une fois le signal transmit on attend ensuite la réponse de l'homme. Si elle ne vient pas après un temps défini on peut éventuellement le relancer à l'aide d'une nouveau signal. Il est également nécessaire de définir quand abandonner la tâche. Un premier facteur est le départ de l'homme mais il n'est pas suffisant, il est aussi utile de considérer son absence de réponse.

En ce qui concerne le *comment* cela dépendra de l'environnement et de la situation actuelle. Par exemple si le robot se trouve dans un environnement qu'il sait bruyant il privilégiera alors les gestes à la parole. S'il est occupé à effectuer à faire une autre action avec ses bras il se contentera alors d'un signe de tête et/ou de la parole.

Enfin, en ce qui concerne l'interprétation des signaux envoyés par l'homme cette partie sera liée à celle gérant l'interprétation de la communication dans la section suivante.

3.2.3 Niveau Motor

Le niveau "motor" est divisé en deux parties principales : une première concernant la gestion des actions du robot et une seconde concernant le monitoring de celles de l'homme. On y retrouvera également la gestion des conflits pouvant exister concernant la tête du robot.

3.2.3.1 Gestion des actions du robot

Dans cette partie on considère trois façons pour le robot d'effectuer une action : l'effectuer seul (seule la bonne exécution de cette action est important), l'effectuer seul mais en présence d'un autre agent (on doit alors produire des signaux concernant cette action afin de la rendre compréhensible) ou alors effectuer une action jointe (action nécessitant l'aide d'un autre agent).

Afin de gérer ces actions une mécanique d'exécution de l'action a été définie. Cette mécanique est la suivante :

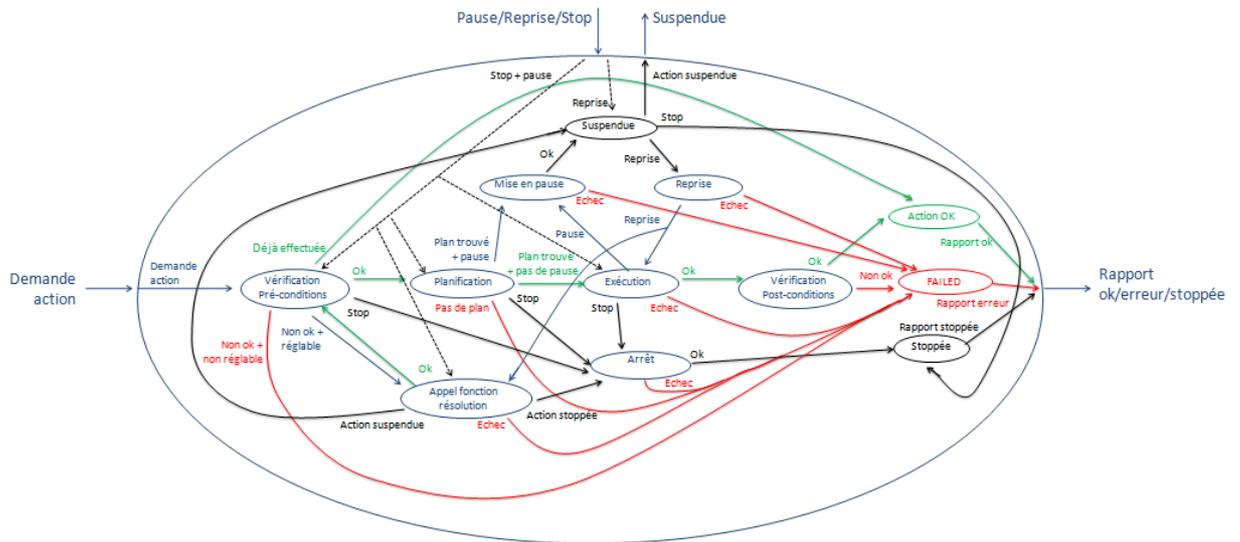


FIGURE 3.2 – Mécanique d'exécution de l'action

Une action est constitué de plusieurs états qui sont :

- *La vérification des pré-conditions* : On vérifie les conditions nécessaires afin d'effectuer l'action (ex : main libre avant de prendre un objet ou alors pince ouverte). Si elles ne sont pas satisfaites il y a deux possibilités : ces pré-conditions ne sont pas facilement corrigibles, l'action échoue alors ou alors elles peuvent l'être, on passera donc dans l'état *appel aux fonctions de résolution*.
- *L'appel aux fonctions de résolution* : Cet état permet, par des actions simples, de fixer une pré-condition nécessaire à l'action (ex : ouvrir une pince avant de prendre un objet).
- *La planification de mouvement* : Dans cet état on produit les trajectoires nécessaires à l'exécution de l'action.
- *L'exécution* : Dans cet état on réalise les actions nécessaires à l'action.

- *La vérification des post-conditions* : On vérifie ou l'on applique les post-conditions de l'action (ex :attachement de l'objet à la main après l'avoir pris).
- *La mise à l'arrêt* : A tout moment un ordre d'arrêt peut arriver de l'extérieur. On sortira alors de l'état en cours pour rentrer dans l'état mis à l'arrêt. Dans cet état, si nécessaire, on arrêtera proprement les mouvements du robot.
- *La mise en pause* : De la même manière que pour l'arrêt, un ordre de pause peut arriver. On devra de la même manière arrêter, si besoin, les mouvements du robot proprement mais également sauvegarder les informations nécessaires à une éventuelle reprise.
- *L'état de suspension* : Dans cet état on attend soit un ordre de reprise, soit d'arrêt définitif.
- *La reprise* : On vérifie si les éléments de l'environnement utiles à l'action n'ont pas été modifiés. Si c'est le cas on peut reprendre là où l'on s'était arrêté, sinon on recommence depuis la vérification des pré-conditions.
- *Les états de sortie* : Il y a donc trois manières de sortir de l'action. Soit l'action s'est déroulée correctement, soit elle a échoué, soit elle a été stoppée. Dans les trois cas on fournira depuis ces états un rapport de sortie concernant le déroulement de l'action.

3.2.3.2 Monitoring de l'homme

Trois fonctions sont nécessaires concernant le monitoring de l'homme, partenaire du robot :

- *Le monitoring des actions effectuées par l'homme* : Ce monitoring est effectué à l'aide des signaux reçus par SPARK. Ces signaux sont analysés en fonction de l'état actuel du monde pour en déduire quelle action a été effectuée (par exemple si la main de l'homme s'approche d'un cube alors qu'il a les mains vides alors on en déduit que l'homme a pris le cube alors que si il avait déjà un autre cube dans la main on peut déduire qu'il a posé ce cube sur le premier).
Une fois l'action effectuée identifiée on va la comparer avec les actions venant du niveau proximal. Si l'action correspond alors on renvoie un rapport comme quoi elle s'est bien déroulée, sinon on renvoie un rapport signifiant une action inattendue.
- *La gestion des interférences avec le robot* : Lors de l'exécution d'une action par le robot il est important de vérifier que les actions de l'homme n'entrent pas en interférence avec celles du robot. Par exemple si le robot est en train de prendre un objet et que l'homme veut attraper cet objet également alors le robot devra arrêter son action. De même si l'homme approche trop près des bras du robot lors d'une action il devra s'arrêter également.
Pour faire cela on utilise deux choses : les faits retournés par SPARK concernant les distances entre l'homme et le robot et les signaux reçus de SPARK concernant le monitoring de l'homme : si l'homme se dirige vers l'objet de l'action du robot alors le robot arrêtera son action.
- *L'interprétation de la communication* : Il est également important de pouvoir interpréter les signaux envoyés par l'homme. Une partie du monitoring de l'homme doit donc être réservée à cet usage. Cependant, pour le moment, par manque de temps et de perception précise de l'homme, cette partie n'a pas été développée.

3.2.3.3 Gestion de la tête du robot

- *Problématique* : La tête du robot est à la fois un élément important concernant la perception du robot mais surtout concernant la communication. En effet, une bonne gestion de là où regarde le robot va beaucoup influencer sur le naturel de l'interaction homme-robot et va servir à faire passer des signaux permettant la communication avec l'homme.
- *Solution implémentée* : Pour aider à la gestion de la tête du robot une gestion à l'aide de queues à priorités était déjà présente dans le superviseur. Les objets ou parties du corps de l'homme à regarder sont rangées dans 3 différentes queues de priorité basse, moyenne ou haute. Le robot regardera ces cibles dans l'ordre de ces priorités jusqu'à ce que les conditions de réussite ou d'échec associée à chaque cible soit validées. Il a donc été choisi d'utiliser pour le moment ces queues de la manière suivante :
 - *Priorité basse* : on rentre dans ce niveau la tête de l'homme, de manière à ce que quand le robot n'a rien de spécial à regarder il regarde l'homme.
 - *Priorité moyenne* : on rentre dans ce niveau toutes les cibles concernant les actions du robot où action jointes (objet de l'action, regard à l'homme pour effectuer des signaux...).
 - *Priorité haute* : on rentre dans ce niveau toutes les cibles concernant les actions de l'homme, qu'elles soit attendues ou non. Ainsi, à chaque fois que l'homme active une sphère de monitoring le robot regardera immédiatement cette dernière afin de montrer son intérêt et de mieux percevoir l'action de l'homme.
- *Améliorations possibles* : Une façon plus fine de décider quoi regarder, quand et également combien de temps peut être envisagée. Il serait également intéressant de pouvoir anticiper légèrement certaines actions de l'homme afin de pouvoir porter une attention plus naturelle à ces dernières.

3.3 Mise en œuvre

L'architecture a été implémentée avec succès et une première version de l'exemple est à présent fonctionnelle. Cette version permet au robot de construire la pile demandée seul où en collaboration avec un homme. Il est capable de se reprendre et de calculer un nouveau plan qu'il exécutera ensuite depuis n'importe quelle situation.

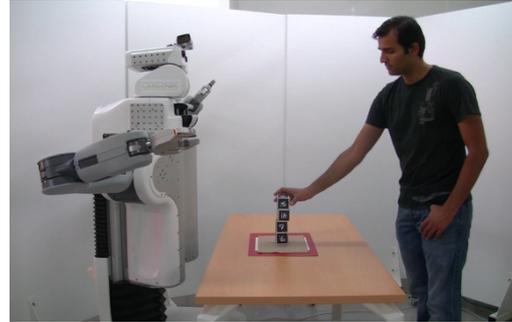
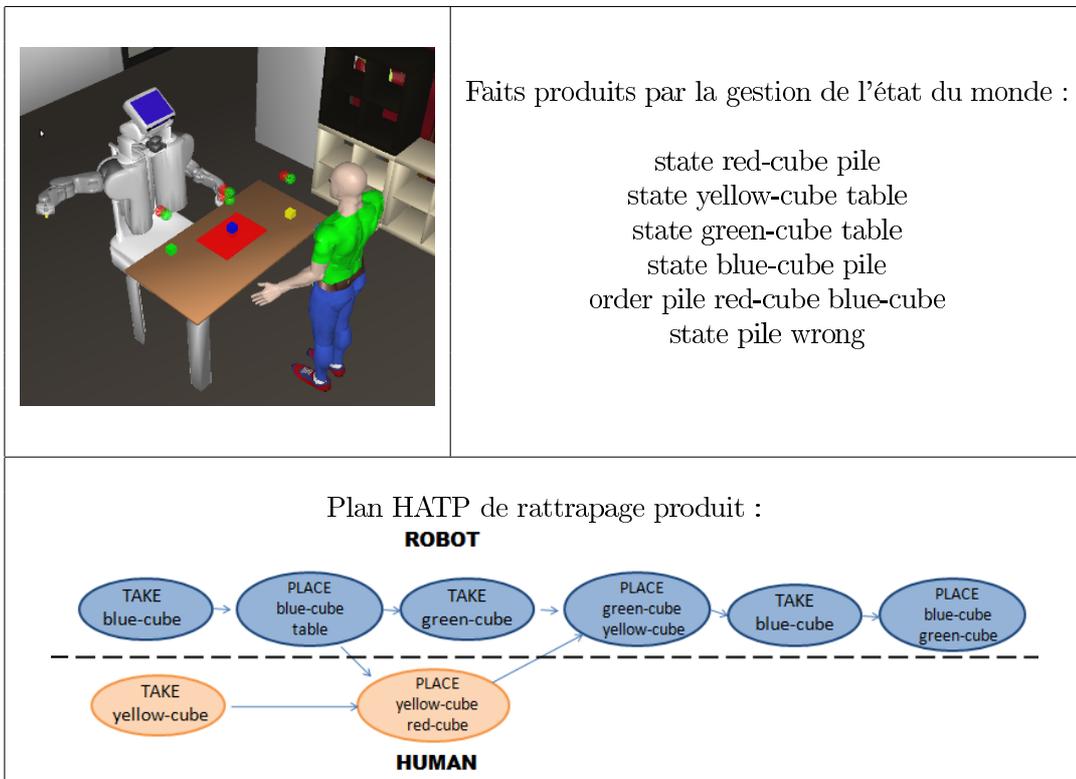


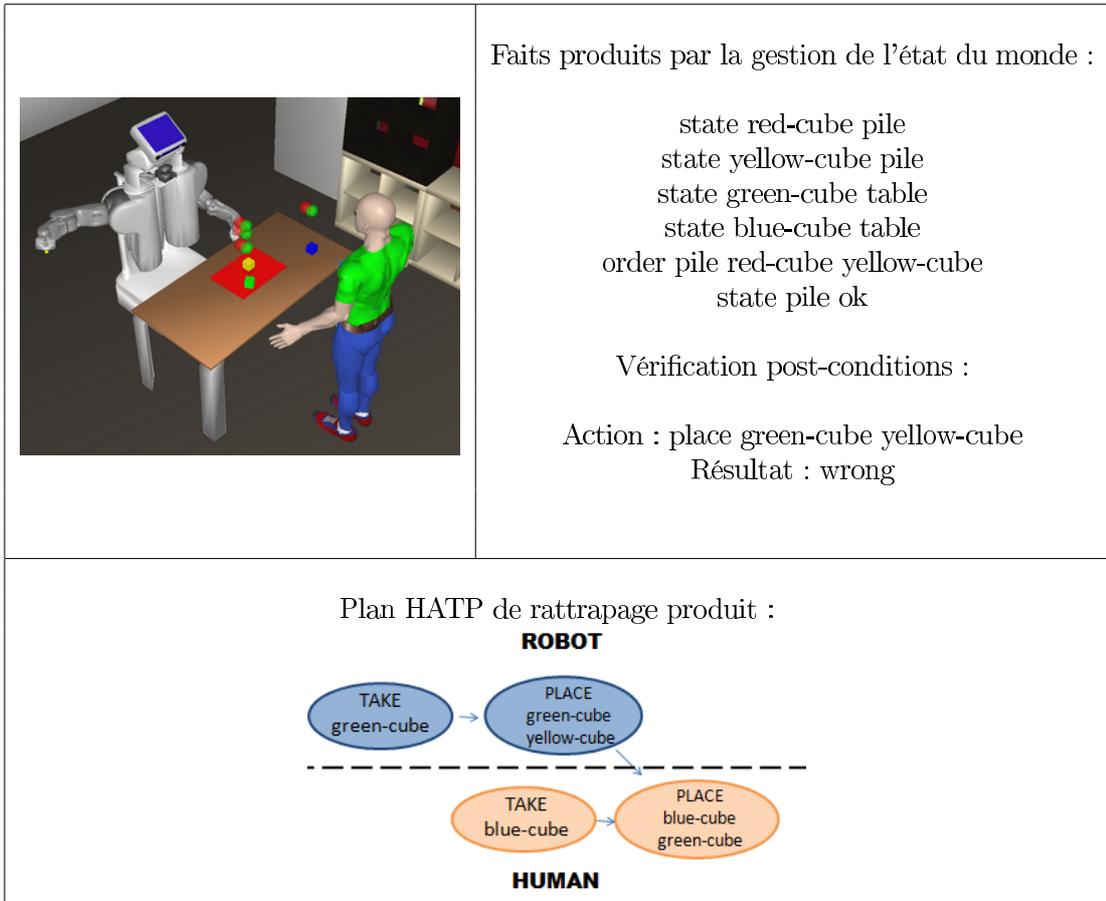
FIGURE 3.3 – Le robot construit la pile seul

FIGURE 3.4 – Le robot et l'homme construisent la pile en collaboration

Le robot est également capable de détecter quand l'homme pose le mauvais cube sur la pile, il calcule alors un nouveau plan permettant la remise de la pile dans un état correct et la fin de sa construction.



Le robot peut également détecter qu'un cube (ou plusieurs) sont tombés de la pile après une action. Il pourra construire un plan pour reconstruire la pile, cependant il n'est actuellement pas forcément en mesure de l'exécuter car les cubes tombés sont souvent bien trop près de la pile pour être attrapés. L'implémentation d'une fonction demandant à l'homme de soit prendre les cubes, soit nettoyer la zone de travail est donc nécessaire pour palier à ce problème. De même, les différentes actions où le robot peut nécessiter l'aide de l'homme (par exemple soutenir la pile pendant qu'il pose un cube si la pile est instable) doivent être définies et la demande d'aide implémentée.



Cependant le système comporte encore plusieurs lenteurs qu'il est nécessaire de corriger.

L'architecture développée peut être utilisée sur d'autre exemple que celui traité. Par exemple, le même superviseur est actuellement utilisé pour préparer une étude utilisateur où le robot doit nettoyer une table en collaboration avec un homme.

Chapitre 4

Perspectives et conclusion

Lors de ce stage j'ai pu mettre en place une architecture permettant la collaboration homme-robot. Cette architecture a également pu être éprouvée sur un exemple concret de travail en collaboration avec un homme.

L'architecture développée étant une base son contenu peut donc être améliorée et enrichie. Quelques moyens de faire cela peuvent être :

- Placer le robot dans un contexte où la (ou les) tâche(s) à effectuer demanderont au robot de se déplacer dans la pièce et de ne plus seulement utiliser ses bras pour agir.
- Placer le robot dans un contexte où il y aurait plus qu'un autre agent (deuxième robot, deuxième homme).
- Approfondir et implémenter la gestion des signaux échangés entre l'homme et le robot.

Cependant, bien que de nombreux points de l'architecture actuelle puissent être améliorés, ce stage a permis de poser de bonnes bases pour la collaboration homme-robot. Une architecture est maintenant disponible grâce à laquelle il est maintenant possible d'implémenter aisément de nouveaux éléments de cette collaboration.

Ce stage m'a également permis d'avoir une vue d'ensemble sur les problèmes liés à l'action jointe homme-robot. Ayant prévu de continuer avec un doctorat sur le même sujet ces connaissances nouvellement acquises me seront je pense très utiles par la suite.

Bibliographie

- [ACF⁺98] R Alami, R Chatila, S Fleury, M Ghallab, and F Ingrand. An architecture for autonomy. *International Journal of Robotics Research*, 17 :315–337, 1998.
- [AWG⁺11] Rachid Alami, Mathieu Warnier, Julien Guitton, Severin Lemaignan, and Emrah Akin Sisbot. When the robot considers the human.. In *Proceedings of the 15th International Symposium on Robotics Research*, pages 1–17, 2011.
- [Bra92] ME Bratman. Shared cooperative activity. *The philosophical review*, 101(2) :327–341, 1992.
- [Bra93] ME Bratman. Shared intention. *Ethics*, 104(1) :97–113, 1993.
- [Bre02] Cynthia Breazeal. Toward sociable robo. *ROBOTICS AND AUTONOMOUS SYSTEMS*, 42, 2002.
- [CAM09] Aurelie Clodic, Rachid Alami, and Vincent Montreuil. Shary : a supervision System Adapted to Human-Robot Interaction. *Experimental Robotics*, 54(Springer Tracts in Advanced Robotics) :229–238, 2009.
- [FCA14] Michelangelo Fiore, Aurelie Clodic, and Rachid Alami. On planning and task achievement modalities for human-robot collaboration. In *International Symposium on Experimental Robotics, Marrakech/Essaouira, June 15–18, 2014*, 2014.
- [FHC97] Sara Fleury, Matthieu Herrb, and Raja Chatila. G eno M : A Tool for the Specification and the Implementation of Operating Modules in a Distributed Robot Architecture 1 Introduction 2 A Network of Modules. In *In International Conference on Intelligent Robots and Systems*, pages 842—848, 1997.
- [FSS⁺06] Terrence Fong, Jean Scholtz, Julie A Shah, Lorenzo Fl, Clayton Kunz, David Lees, John Schreiner, Michael Siegel, Laura M Hiatt, Illah Nourbakhsh, Reid Simmons, Robert Ambrose, Robert Burrige, Brian Antonishek, Magda Bugajska, Alan Schultz, J Gregory Trafton, and A The Peer-to-peer Human-robot Interaction Project. A Preliminary Study of Peer-to-Peer Human-Robot Interaction. In *IEEE International Conf. on Systems, Man, and Cybernetics*, 2006.
- [GBB⁺05] J. Gray, C. Breazeal, M. Berlin, A. Brooks, and J. Lieberman. Action parsing and goal inference using self as simulator. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication*, pages 202–209. Ieee, 2005.
- [Gib79] James J. Gibson. The Theory of Affordances. In *The Ecological Approach to Visual Perception*, pages 127–137. 1979.

- [HB04] Guy Hoffman and Cynthia Breazeal. Robots that work in collaboration with people. In *AAAI Fall Symposium on the Intersection of Cognitive Science and Robotics*, pages 1–5, 2004.
- [Hof10] Guy Hoffman. Anticipation in Human-Robot Interaction. In *AAAI Spring Symposium : It's All in the Timing*, pages 21–26, 2010.
- [KBS11] G Knoblich, S Butterfill, and Natalie Sebanz. Psychological Research on Joint Action : Theory and Data. *The Psychology of Learning and Motivation*, 54 :59–101, 2011.
- [KWB⁺04] G. Klein, D.D. Woods, J.M. Bradshaw, R.R. Hoffman, and P.J. Feltovich. Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems*, 19(6) :91–95, 2004.
- [MWCA14] Gregoire Millez, Matthieu Warnier, Aurelie Clodic, and Rachid Alami. A framework for endowing an interactive robot with reasoning capabilities about perspective-taking and belief management. In *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication*, 2014.
- [Pac08] Elisabeth Pacherie. The phenomenology of action : a conceptual framework. *Cognition*, 107(1) :179–217, April 2008.
- [Pac12] Elisabeth Pacherie. The phenomenology of joint action : self-agency vs. joint-agency. *Joint attention : New developments*, 93 :343–389, 2012.
- [SBK06] Natalie Sebanz, Harold Bekkering, and Günther Knoblich. Joint action : bodies and minds moving together. *Trends in cognitive sciences*, 10(2) :70–76, February 2006.
- [SK09] Natalie Sebanz and Guenther Knoblich. Prediction in Joint Action : What, When, and Where. *Topics in Cognitive Science*, 1(2) :353–367, April 2009.
- [SLL01] T Simeon, J-p Laumond, and F Lamiroux. Move3D : a generic platform for path planning. in *4th Int. Symp. on Assembly and Task Planning*, pages 25–30, 2001.
- [TCC⁺05] Michael Tomasello, Malinda Carpenter, Josep Call, Tanya Behne, and Henrike Moll. Understanding and sharing intentions : the origins of cultural cognition. *The Behavioral and brain sciences*, 28(5) :675–691, October 2005.
- [VBKS10] Cordula Vesper, Stephen Butterfill, Günther Knoblich, and Natalie Sebanz. A minimal architecture for joint action. *Neural networks : the official journal of the International Neural Network Society*, 23(8-9) :998–1003, 2010.
- [WGLA12] Mathieu Warnier, Julien Guitton, Severin Lemaignan, and Rachid Alami. When the robot puts itself in your shoes. Managing and exploiting human and robot beliefs. In *2012 IEEE RO-MAN : The 21st IEEE International Symposium on Robot and Human Interactive Communication*, pages 948–954. Ieee, September 2012.