When the robot considers the human

Rachid Alami, Mathieu Warnier, Julien Guitton, Séverin Lemaignan, Emrah Akin Sisbot

When the robot considers the human...

Rachid Alami, Mathieu Warnier, Julien Guitton, Séverin Lemaignan and Emrah Akin Sisbot

Abstract This paper addresses some key decisional issues that are necessary for a cognitive robot which shares space and tasks with a human. We adopt a constructive approach based on the identification and the effective implementation of individual and collaborative skills. The system is comprehensive since it aims at dealing with a complete set of abilities articulated so that the robot controller is effectively able to conduct in a flexible manner a collaborative task with a human partner. 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 planning and human and robot interleaved plan achievement.

1 Introduction

Human-robot interaction requires to equip the robot with explicit reasoning on the human and on its own capacities to achieve its tasks in a collaborative way with a human partner. This paper presents a robot control system which has been especially designed for a cognitive robot which shares space and task with a human. We have adopted a constructive approach based on effective individual and collaborative skills. The system is comprehensive since it aims at dealing with a complete set of abilities articulated so that the robot controller is effectively able to conduct in a flexible manner a collaborative task with a human partner.

We illustrate below how we deal with a typical human-robot interactive task achievement and what are the abilities we claim are necessary. Section §2 proposes a typical human-robot interaction problem that can be solved by the proposed robot

LAAS-CNRS
Université de Toulouse, UPS, INSA, INP, ISAE, LAAS
F-31077 Toulouse, France
e-mail: Firstname.Name@laas.fr

controller. Section §3 reviews related work and analyses the context of our contribution. Section §4 provides an overview of the robot controller and introduces three activities which are described in §5, §6 and §7. Finally, Section §8 presents an effective run on a real robot in face to face interaction with a person.

2 A typical human-robot interaction problem

Let us consider a robot which is supposed to achieve interactive object manipulation, fetch and carry tasks and similar tasks in a domestic environment. The problem we are dealing with here is the following. Given:

- a joint goal, which has been previously established and agreed upon (through a process which is out of the scope of this paper),
- the current situation, acquired through perception or deduction from previous perceptions, including the state of the environment of the robot and of the human,

the robot controller computes an action to execute and who (the human or the robot, or both in case of a joint action) has to perform it, and then controls or monitors its execution. The operation continues until the goal is achieved, is declared unachievable or is abandoned by the human.

To do so, the robot has to be equipped with a number of decisional, planning and interpretation abilities where its human partner is taken explicitly into account. It needs to be able:

- to build and maintain relevant robot and human beliefs (from the robot perspective) with respect to state of the world and the task,
- to build and maintain iteratively shared (human-robot) plans,
- to refine and execute the actions it has to perform, and to monitor those achieved by its human partner.

Besides, we would like to build such abilities in a generic way, and to provide several levels of parametrization allowing to adapt to various environments, and various levels of involvement of the robot ranging from teammate behavior to assistant or proactive helper.

3 Related work and vision

The human presence brings new requirements for robot's abilities both at the functional and at the deliberative levels [17]. The topics involve motion [19, 5, 22], navigation [3, 33], manipulation [16] in presence of humans as well as perception of human activities [8, 9]. Also, when interacting with humans, robots need to incorporate communication and collaboration abilities. Several theories dealing with collaboration [11, 14, 10] emphasize that collaborative tasks have specific requirements compared to individual ones, *e.g.* , since the robot and the person share a common goal, they have to agree on the manner to realize it, they must show their commitment to the goal during execution, etc. Several robotic systems have already been built based on these theories [28, 31, 36, 6] and they all have shown benefits of this approach. They have also shown how difficult it is to manage turn-taking between communication partners and to interleave task realization and communication in a generic way. Finally, today only few systems [13, 6, 32] take humans into account at all levels.

Perspective Taking is a human ability which allows one to put him/herself in another person's point of view. Studied in psychology literature [12, 39], this ability is crucial when interacting with people by allowing one to reason on others' understanding of the world in terms of visual perception, spatial descriptions, affordances and beliefs, etc. Therefore, in the last years these notions have been gradually employed in Human-Robot Interaction. [7] presents a learning algorithm that takes into account information about a teacher's visual perspective in order to learn a task. [15] apply visual perspective taking for action recognition between two robots. [38] use both visual and spatial perspective taking for finding out the referent indicated by a human partner.

Spatial reasoning [26], on the other hand, has been used for natural language processing for applications such as direction recognition [18, 24] or language grounding [37]. [35] presented a spatial reasoner integrated in a robot which computes symbolic positions of objects.

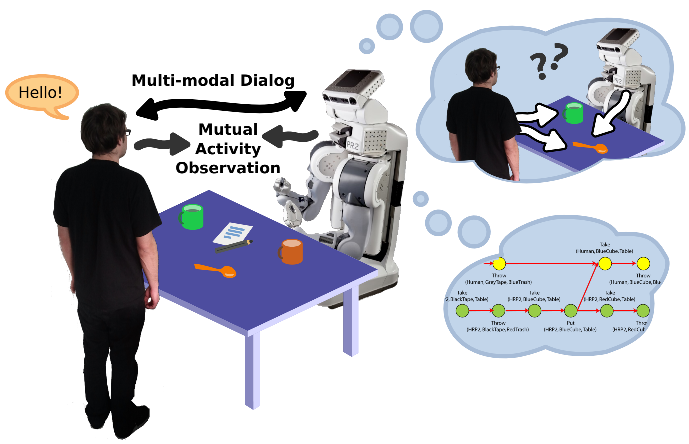


Fig. 1 Robot reasoning about HRI and anticipation of human activities: sources of information are multi-modal dialogue, and observation of environment and human activity

We envision HRI in a context where two agents (a human and a robot) share a common space and exchange information through various modalities. Our aim is to endow the robot with an explicit consideration of the human and with the ability to manage its interactions with him (Figure 1). This must be considered at the architecture level as well as at the task/motion planning and execution level.

We have devised a decisional framework for human-robot interactive task achievement that is aimed to allow the robot not only to accomplish its tasks but also to produce behaviors that support its engagement vis-a-vis its human partner and to interpret human behaviors and intentions. Together and in coherence with this framework, we have developed and experimented various task planners and interaction schemes that allow the robot to select and perform its tasks while taking into account explicitly the human abilities as well as the constraints imposed by the presence of humans, their needs and preferences.

4 A Decisional framework

Interaction happens as a consequence of an explicit request of the human to satisfy a goal or because the robot finds itself in a situation where it is useful if not mandatory. In both cases, the robot has a goal to satisfy. An important issue is the notion of engagement, a process in which the robot will have to establish, maintain and terminate a connection with a human partner. This covers goal establishment, selection of an incremental refinement of the task that is intended to be achieved, and execution control including monitoring, and even influencing, human task performance and his/her commitment to the goal. The human involvement may range from a direct participation to the task achievement, to a simple "acceptance" of robot activity in his/her close vicinity.

Our robot is controlled by a three layer architecture [1]. We present briefly its decisional layer. The proposed decisional framework consists of several entities, having each a specific role as illustrated by Figure 2. We describe how the robot is controlled through an analysis of the three main activities performed by the robot controller:

1. Situation assessment and context management
2. Goals and plans management
3. Action refinement, execution and monitoring

The next three sections describe the three robot controller activities and how they make use of a number of key components in the architecture:

- SPARK: Spatial Reasoning and Knowledge module [34]
- ORO: a knowledge management module [20]
- HATP: a Human-Aware Task Planner [2]
- A set of Human aware motion, placement and manipulation planners [32, 23, 27]

Other decisional activities, such as situated dialog ([29, 21], not presented here) have been developed that use the same set of components.

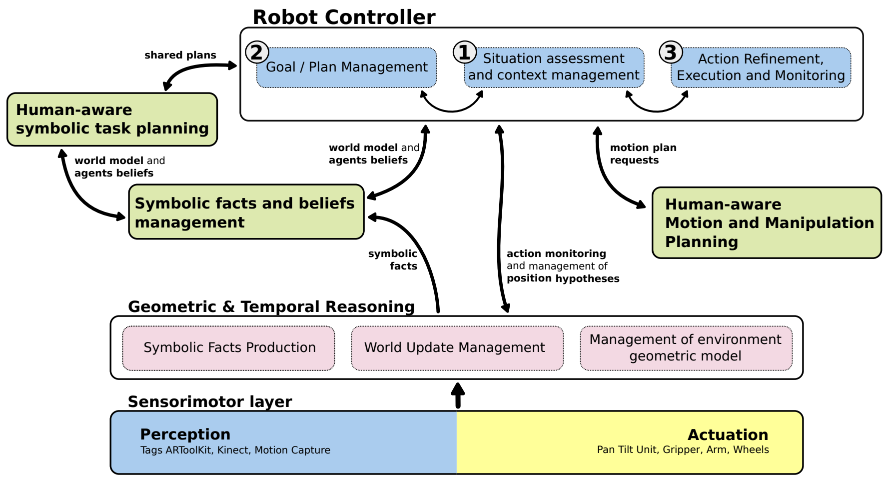


Fig. 2 Architecture of the robot control system

5 Situation assessment and context management

This activity involves the geometric and temporal reasoning component, the symbolic facts and belief management component and the dedicated robot controller activity (Figure 2).

We assume that perception provides in real-time the identity and the position of objects when they are in the field of view of the sensors. In our implemented examples, the robot is localised using a standard horizontal laser-scanning based localisation system, the objects are identified and localized using ARToolkit [4] and the humans are tracked using a commercial motion capture system and a Kinect device from Microsoft.

5.1 Geometric and Temporal Reasoning component

The geometric reasoning component plays a central role in our architecture. It is called SPARK (Spatial Reasoning and Knowledge [34]) in the current implementation. It is responsible for geometric information gathering and it embeds a number of decisional activities linked to abstraction (symbolic facts production) and inference based on geometric and temporal reasoning. SPARK maintains all geometric positions and configurations of agents, objects and furniture coming from perception and previous or *a priori* knowledge.

Symbolic facts production: Geometric state of the world is abstracted in symbolic facts that can be classified in three different categories.

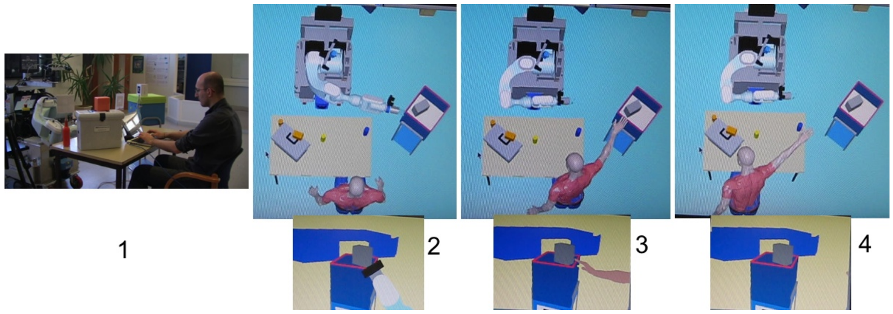


Fig. 3 An example illustrating the *reachable* relation. The relation is computed from the perspectives of both the robot and the human. The computed posture at each step is illustrated with a global view of the scene (top), and from a closest view (bottom).

- Relative positions of object and agents, e.g. ⟨GREY_TAPE isOn TABLE⟩.
- Perception and manipulation capacity and state of agents, e.g. ⟨ROBOT looksAt GREY_TAPE⟩, ⟨GREY_TAPE isVisibleBy HUMAN1⟩.
- Motion status for object or agent parts, e.g. ⟨GREY_TAPE isMoving true⟩, ⟨ROBOT_HEAD isTurning true⟩.

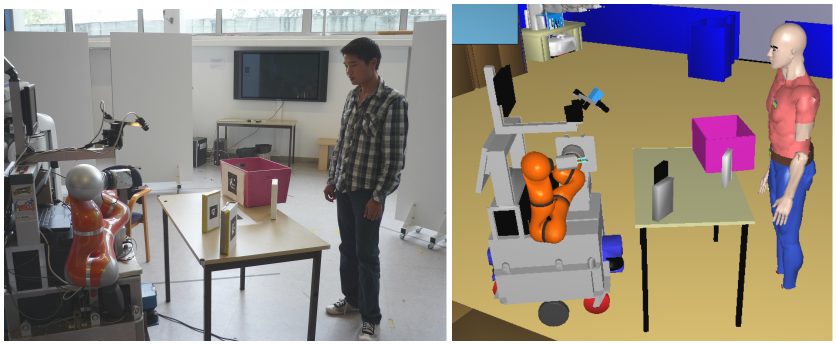
Reasoning about human perspective allow to compute facts such as: ⟨GREY_TAPE isBehind HUMAN1⟩, ⟨GREY_TAPE isVisibleBy HUMAN1⟩.

Figure 3 illustrates different situations for the *reachable* relation. In this case, the robot and its human partner are placed face to face, in a table-top setup (Figure 3.1). The robot first estimates if the small grey box is reachable to itself. This is done by finding a collision free posture to reach the object (Figure 3.2). Next the robot switches to the human's perspective to estimate if the same object is reachable to the human as well. In the last scene, the human moves towards his left, farther from the object (Figure 3.4). The situation is then reevaluated. In this occasion though, the reasoner cannot find a satisfactory posture for the human to reach the box because he is too far from the target.

The set of facts computed in the situation depicted by Figure 5.1 is the following:

ROBOT	HUMAN1
PINK_TRASHBIN isReachable false	PINK_TRASHBIN isReachable true
WALLE_TAPE isReachable false	WALLE_TAPE isVisible true
LOTR_TAPE isReachable true	LOTR_TAPE isReachable false
GREY_TAPE isReachable true	GREY_TAPE isReachable false
WALLE_TAPE isVisible true	WALLE_TAPE isReachable true
LOTR_TAPE isVisible true	LOTR_TAPE isVisible true
GREY_TAPE isVisible true	GREY_TAPE isVisible true
WALLE_TAPE isOn TABLE	WALLE_TAPE isOn TABLE
LOTR_TAPE isOn TABLE	LOTR_TAPE isOn TABLE
GREY_TAPE isOn TABLE	GREY_TAPE isOn TABLE

Hypotheses on objects states and positions: It is sometimes difficult or even impossible to see and/or track an object in certain states. This happens, for instance,



(a) Initial state (b) 3d model view of initial state

Fig. 4 In this situation, there are three tapes on the table. Two tapes are only reachable by the robot: the LOTR_TAPE (black in the 3d model) and GREY_TAPE. The third tape WALLE_TAPE (white in the 3d model) and the trashbin PINK_TRASHBIN are only reachable by the human HUMAN1. All tapes are on the table TABLE.

when the object has been put in a container, when it is in the robot gripper or in the human hand, and more generally in any state in which it is hidden by something else. Our robot has a model of the possible symbolic states for an object (whether the object is on a furniture, in an agent hand, in a container, etc.). According to the robot perception of what has happened since the object was last seen, the robot tries to maintain a belief of the current possible symbolic states and their associated probabilities for this object. Such information can be used to update the beliefs using input from exploration, dialog, human visual focus,. . . SPARK currently provides a simple implementation of such a functionality. The only managed hypotheses are *in container* and *in agent hand*. We can have only one hypothesis at the same time. Hypothesis validity is checked geometrically in case of incoming perception values.

Primitive action recognition: Monitoring human activity is crucial to maintain a coherent state of the world. Full human action and activity monitoring is a difficult task that requires knowledge and reasoning both on high level facts like goals, intentions and plans, as well as bottom-up data from agent and object motions. Simple temporal and geometric reasoning on human hand trajectories and potential objects placements can provide some useful clues for high level human monitoring processes. We call this temporal and geometric reasoning *primitive action recognition*.

For example, a *pick*, a *throw* or a *place* action can be recognized by observing that an object on table and an empty human hand are close to each other, or that the human hand holding an object is close to a container, etc. Human hand position is either directly perceived or inferred from its initial perceived trajectory. We have a simple implementation of such a primitive action recognition in SPARK that relies on monitoring human hand and its motion near objects or above containers.

5.2 Symbolic facts and beliefs management

The facts produced by the geometric and temporal reasoning component are stored in a central symbolic knowledge base, called ORO. Besides acting as a facts database, the ORO platform [20] exposes several functions: operations on knowledge statements relying on inference (through a continuous first-order logic classification process), management of *per-agent* symbolic models, and also higher cognitive and human-robot interaction related functionalities like categorization of sets of concepts, profiles of memory (that enable the robot to "forget" about some facts), natural language grounding [21]. . . .

ORO stores independent knowledge models (in our implementation, as *ontologies*) for each agent (the robot and the humans it interacts with). The robot architecture components (like the executive layer or the situation assessment component) can then store the agents' beliefs in specific models. Each of these models is independent and logically consistent, enabling reasoning on different perspectives of the world that would otherwise be considered as globally inconsistent (for instance, an object can be visible for the robot but not for the human. This object can have at the same time the property `isVisible` **`true`** and `isVisible` **`false`** in two different models). This feature actually allows us to consider the robot to be endowed with a simple *theory of mind* [30]: it can explicitly model the belief state of its interactors.

ORO also provides an event mechanism that allows components to be triggered when specific events occur. A component can for instance subscribe to events of kind `[?agent isVisible true, ?agent type Human]`. As soon as the perception layer detects a human in the robot's field of view and accordingly updates the knowledge base, the executive layer would be triggered back. The event framework also takes advantage of the inference capabilities of ORO. Thus an event can be indirectly triggered if its triggering conditions can be inferred to be true.

5.3 Situation Assessment and Context Management Controller

Building, updating and maintaining a correct state of the world at geometric and symbolic level is crucial to the capacity of the robot to carry on successfully a multi-step interaction with a human. Tight integration between the robot controller and the geometric and temporal reasoning functions in SPARK and symbolic facts and beliefs management in ORO is central.

The robot controller has access to the symbolic facts in ORO that are automatically updated whenever object and agent positions are changed. Robot controller can also access geometric perceived or inferred positions of objects and geometric positions and postures of the human that will be used to orient its cameras. Building and updating the state of the world first relies on perceiving objects. Robot controller can use:

- Exploration policies: robot will exhaustively scan the table to see all what can be seen.
- Search policies: robot will search an object until it is detected if possible, scanning all the table and looking in human hand.

Robot reasons on possible positions for non perceived objects. These hypotheses could be updated using new input from dialogue, human action and focus of attention. Currently, we manage at most one hypothesis per object. This hypothesis is produced by robot controller through an inference on robot or human action. In case of perception conflicts with low probability for the current hypothesis, robot controller will break this hypothesis and delete corresponding symbolic fact in the ontology.

Robot must reacts to change in the world not linked to robot action to drive world update. Robot controller uses SPARK to monitor human hand motion and primitive action recognition for *pick* and *throw*. As mentioned above these primitive action recognition should be used with higher level information on goals, intentions and plans and some exploration to achieve complex human action and activity monitoring. In the current implementation, these primitive actions are over-optimistically interpreted as the corresponding actions *pick* object and *throw* object.

6 Goal and Plan Management

The Goal and Plan Management activity involves the human-aware symbolic task planner component and the dedicated robot controller activity (Figure 2).

6.1 Symbolic Task Planning

In order to devise how a given goal can be accomplished, the robot has to elaborate a plan, *i.e.* a set of actions to be achieved by the robot and its human partners. This is the role of a HATP [2] (for Human Aware Task Planner). HATP is based on a Hierarchical Task Network (HTN) refinement which performs an iterative task decomposition into sub-tasks until reaching atomic actions [25]. The planning domain defines a set of methods describing how to decompose a task and can be seen as the Howto knowledge of the robot. HATP is able to produce plans for the robot's actions as well as for the other participants (humans or robots). It can be tuned by setting up different costs depending on the actions to apply and by taking into account a set of constraints called social rules. This tuning aims at adapting the robot's behavior according to the desired level of cooperation of the robot.

Agents and action streams: The robot plans not only for itself but also for the other agents. The resulting plan, called "shared plan" is a set of actions that form a stream for each agent involved in the goal achievement. Depending on the context, some "shared plans" contain causal relations between agents. For example, the second agent needs to wait for the success of the first agent's action to be able to start its own action. When the plan is performed, causal links induce synchronization between agents. Figure 5 illustrates a plan with two streams.

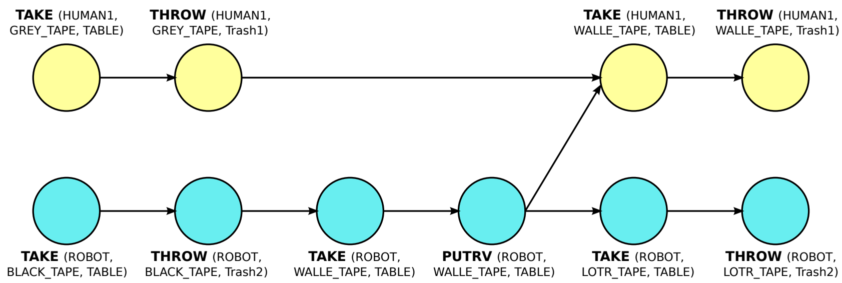


Fig. 5 A plan produced by HATP with 2 streams

Action costs and social rules: A cost and a duration function is associated to each action. The duration function provides a duration interval for the action achievement and is used, in one hand, to schedule the different streams and, in the other hand, as an additional cost function. In addition to these costs, HATP also takes into account a set of social rules. Social rules are constraints aiming at leading the plan construction towards the best plan according to some human preferences. The social rules we have defined so far deal with:

- undesirable state: to avoid a state in which the human could feel uncomfortable;
- undesirable sequence: to eliminate sequences of actions that can be misinterpreted by the human;
- effort balancing: to adjust the work effort of the agents;
- wasted time: used to avoid long delays between the actions of the human partner;
- intricate links: to limit dependencies between the actions of two or more agents.

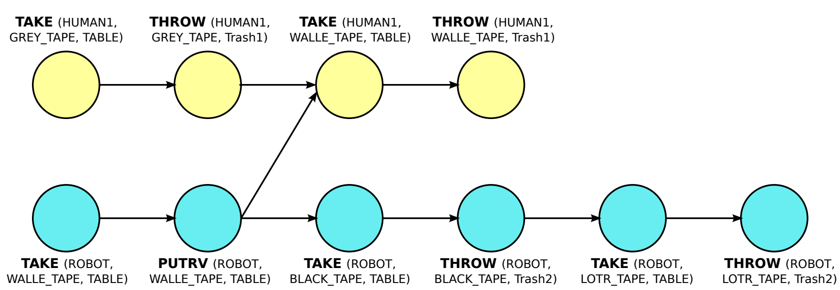


Fig. 6 A plan with the wasted time social rule

Figure 6 illustrates an alternative plan to the previous one (Figure 5) if the wasted time social rule is used. The obtained shared plan is the best plan according to a global evaluation of these multiple criteria.

Several levels of cooperation: By tuning its costs and adapting its social rules, HATP can be used to compute various alternative plans. These plans can be categorized into several levels of cooperation

- helping the human to achieve his goal by acting for him
- sharing concrete resources by handing some objects
- collaboration of the robot and the human by coordinating their actions towards a human-robot joint goal.

6.2 Goal and plan Controller

Figure 7 sums up the Goal and Plan Management as implemented in the robot controller. When an event announcing a new goal is caught by the controller, the validity of this goal is tested: does it corresponds to capabilities of agents? is it not already achieved? Then, the goal is sent to HATP which produces a first plan. A goal is considered achievable as long as the planner computes a valid plan and it is not abandoned by the human.

Plan execution consists in the management of all the actions of the plan. Human and robot are not acting both at the same time. In case of plan failure a new plan is requested and executed.

The management at the action level is done in three steps. First, the action preconditions are tested over the current state of the world. Then the action is executed and monitored (only monitored of it corresponds to a human action). Finally, the expected effects are verified in order to acknowledge the action achievement.

Concerning the speech acts during plan management, the robot informs its partners on the goal existence and status, plan existence and status, ongoing plan action, plan action failure, failing facts for action precondition or effect assessment.

7 Action execution and monitoring

The action execution and monitoring task involves the motion and task planning, the effectors and the dedicated robot controller activity (Figure 2).

Human-aware motion, placement and manipulation planning: In our scenarios, actions are only object manipulation actions: Pick, Place, Throw, Give. Motion and Task planning allows to compute final object placement, grasp and arm motion trajectories taking into account task specific constraints and human postures, abilities and preferences: see [32, 23, 27] for details.

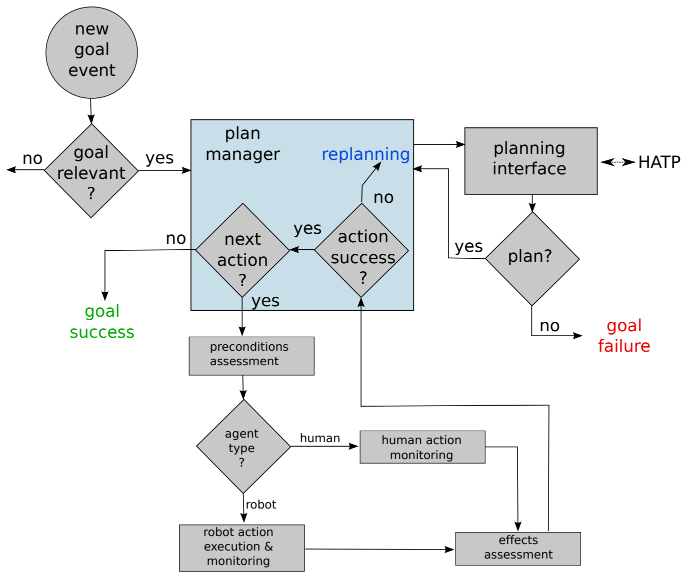


Fig. 7 Automaton for goal and plan management

Execution and Monitoring Robot Controller: Depending on the context and on the shared plan elaborated by HATP for a given goal, the robot controller decides to execute an action or to ask its human partner to do it. Actions feasibility by the human or the robot are regularly reconsidered based on the reachability / visibility computation mechanisms.

Robot action execution is based on simple automatons that translate symbolic planning atomic actions into sequences of planned arm motions and gripper commands to execute according to current state of the 3-tuple (gripper, object, furniture). We have three states according to whether the object is in gripper and if it is in gripper whether it is on furniture. These states are directly obtained from the updated symbolic state of the world in the ontology.

For plan action monitoring, primitive actions recognition defined above are used. A primitive action detection is interpreted as action success if it is the expected one and failure otherwise. The robot also reacts to the absence of activity.

8 An illustrative example

We assume here that the robot (and the human) has been given the joint goal “CLEAN TABLE”. For HATP, this means putting all tapes that are currently on the table in the trashbin. Depending on the state of the world and agent preferences, different plans are produced.

Figure 8 shows a plan produced to clean the table based on the initial the initial state given in § 5.1.

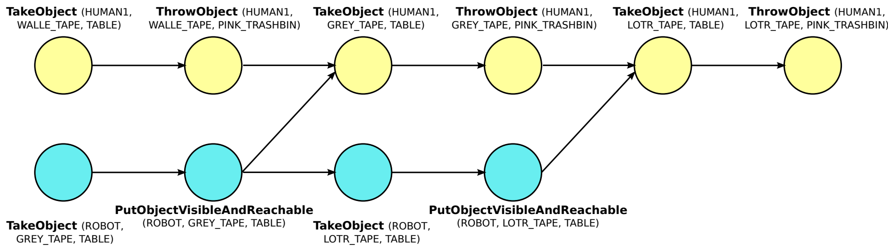


Fig. 8 A plan produced by HATP for clean the table based on the initial state given in § 5.1

Let us now take a simpler example to illustrate a full run of the system. We have only one tape on the table and it is is reachable only by the robot while the throw position on top of the trashbin is reachable only by the human.

Figure 8 illustrates the main processes occurring during a multi-step human-robot collaborative goal achievement. The plan produced is quite straightforward and is shown in the third row called "Goal and Plan". It consists in 4 successive actions involving the robot and the human. Robot grasps the tape and then places it on the table at a position where it is visible and reachable for the human. Human then is asked to pick the tape and throw it in the trashbin. The first row, named "Cameras", shows several snapshots corresponding to various execution steps. Snapshot 1 presents the initial situation. Snapshots 2, 3, 4 and 5 give the state after the successive achievement of the four actions in the plan. The second row, named "3D Model", shows the display of SPARK at the same instants. The fourth, row called "Robot Speech Acts", gives robot speech acts produced along the execution to inform the human partner about goal and plan creation and status and to verbalize the actions that the human is asked to execute. The fifth row illustrates robot knowledge on itself and on the objects. The sixth row illustrates the robot knowledge about the human state. The seventh row gives ongoing robot action with action preconditions and effects assessment as well as motion execution tasks. Motion trajectory typology can be found between the 3D Model views. The eighth row gives ongoing human action with action preconditions and effects assessment and monitoring activity.

9 Conclusion and future work

In this paper we have presented a decisional framework designed for robots operating in a human environment. Our objective is to provide a management of human-robot interaction that is an integral part of a general robot control architecture. This was done in order to provide a principled way to deal with HRI scenario. The frame-

Fig. 9 An example of a human-robot collaborative goal achievement

work is also suitable for the development and experiment of task planners and interaction schemes that explicitly consider human abilities and preferences.

The design choices and the results presented here are still preliminary. While the general scheme we propose might be difficult to implement in a general sense, we believe that it is a reasonable challenge to implement it in the case of a personal robot assistant essentially devoted to fetch-and-carry, as well as for interactive manipulation tasks and associated activities.

Acknowledgements: This work has been conducted within the EU CHRIS project (`http://www.chrisfp7.eu/`) funded by the E.C. Division FP7-IST under Contract 215805. We also would like to thank warmly a number of LAAS robotics team members and more particularly Xavier Broquere, Mokhtar Gharbi, Wuwei He, Matthieu Herrb, Jim Mainprice, Raquel Ros, Daniel Sidobre, Thierry Siméon, Muhammad Ali.

References

1. R. Alami, R. Chatila, S. Fleury, M. Ghallab, and F. Ingrand. An architecture for autonomy. *International Journal of robotics Research, Special Issue on Integrated Architectures for Robot Control and Programming*, 17(4), 1998.
2. S. Alili, V. Montreuil, and R. Alami. HATP task planner for social behavior control in autonomous robotic systems for hri. In *The 9th International Symposium on Distributed Autonomous Robotic Systems*, 2008.
3. P. Althaus, H. Ishiguro, T. Kanda, T. Miyashita, and H.I. Christensen. Navigation for human-robot interaction tasks. *IEEE Int. Conf. on Robotics & Automation, New Orleans, USA*, 2004.
4. ARToolkit. Documentation and download site. *http://www.hitl.washington.edu/artoolkit/*.
5. J. van den Berg and M. Overmars. Roadmap-based motion planning in dynamic environments. Technical report, Utrecht University, NL, 2004.
6. C. Breazeal. Towards sociable robots. *Robotics and Autonomous Syst.*, pages 167–175, 2003.
7. C. Breazeal, M. Berlin, A. Brooks, J. Gray, and A. Thomaz. Using perspective taking to learn from ambiguous demonstrations. *Robotics and Autonomous Systems*, pages 385–393, 2006.
8. C. Breazeal, A. Edsinger, P. Fitzpatrick, and B. Scassellati. Active vision for sociable robots. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 31(5):443–453, 2001.
9. B. Burger, I. Ferrane, and F. Lerasle. Multimodal interaction abilities for a robot companion. In *ICVS*, pages 549–558, 2008.
10. H. H. Clark. *Using Language*. Cambridge University Press, 1996.
11. P. R. Cohen and H. J. Levesque. Teamwork. *Nous*, 25(4):487–512, 1991.
12. J. Flavell. *Perspectives on Perspective Taking*, pages 107–139. L. Erlbaum Associates, 1992.
13. T. W. Fong, C. Kunz, L. Hiatt, and M. Bugajska. The human-robot interaction operating system. In *2006 Human-Robot Interaction Conference*. ACM, March 2006.
14. B. J. Grosz and S. Kraus. Collaborative plans for complex group action. *Artificial Intelligence*, 86:269–358, 1996.
15. M. Johnson and Y. Demiris. Perceptual perspective taking and action recognition. *Advanced Robotic Systems*, 2(4):301–308, 2005.
16. C.C. Kemp and E. Edsinger, A. Torres-Jara. Challenges for robot manipulation in human environments. *Robotics & Automation Magazine, IEEE*, 2007.
17. G. Klein, J. M. Woods, D. D.and Bradshaw, and P. J. Hoffman, R. R.and Feltovich. Ten challenges for making automation a ”team player” in joint human-agent activity. *IEEE Intelligent Systems*, 19(6):91–95, 2004.

18. T. Kollar, S. Tellex, D. Roy, and N. Roy. Toward understanding natural language directions. In *HRI*, pages 259–266, 2010.
19. D. Kulic and E. Croft. Pre-collision safety strategies for human-robot interaction. *Autonomous Robots*, 22(2):149–164, 2007.
20. S. Lemaignan, R. Ros, L. Mosenlechner, R. Alami, and M. Beetz. Oro, a knowledge management module for cognitive architectures in robotics. In *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010.
21. S. Lemaignan, A. Sisbot, and R. Alami. Anchoring interaction through symbolic knowledge. In *Proceedings of the 2011 Human-Robot Interaction Pioneers workshop*, 2011.
22. K. Madhava, R. Alami, and T. Simeon. Safe proactive plans and their execution. *Robotics and Autonomous Systems*, 54(3):244–255, 2006.
23. J. Mainprice, E.A. Sisbot, L. Jaillet, J. Cortes, R. Alami, and T. Simeon. Planning human-aware motions using a sampling-based costmap planner. In *IEEE International Conference on Robotics and Automation*, 2011.
24. C. Matuszek, D. Fox, and K. Koscher. Following directions using statistical machine translation. In *Int. Conf. on Human-Robot Interaction*. ACM Press, 2010.
25. D. Nau, T. C. Au, O. Ilghami, U. Kuter, J. W. Murdock, D. Wu, and F. Yaman. Shop2: An htn planning system. *Journal of Artificial Intelligence Research*, pages 379–404, 2003.
26. J. O'Keefe. *The Spatial Prepositions*. MIT Press, 1999.
27. A.K. Pandey and R. Alami. Mightability maps: A perceptual level decisional framework for co-operative and competitive human-robot interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010.
28. C. Rich and C. L. Sidner. Collagen: When agents collaborate with people. *Proceedings of the first international conference on Autonomous Agents*, 1997.
29. R. Ros, S. Lemaignan, E.A. Sisbot, R. Alami, J. Steinwender, K. Hamann, and F. Warneken. Which one? grounding the referent based on efficient human-robot interaction. In *19th IEEE International Symposium in Robot and Human Interactive Communication*, 2010.
30. B. Scassellati. Theory of mind for a humanoid robot. *Autonomous Robots*, 12(1):13–24, 2002.
31. C. L. Sidner, C. Lee, C. Kidd, N. Lesh, and C. Rich. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1-2):140–164, 2005.
32. E. A. Sisbot, A. Clodic, R. Alami, and M. Ransan. Supervision and motion planning for a mobile manipulator interacting with humans. 2008.
33. E.A. Sisbot, L.F. Marin-Urias, R. Alami, and T. Simeon. A human aware mobile robot motion planner. *IEEE Transactions on Robotics*, 23(5):874–883, 2007.
34. E.A. Sisbot, R. Ros, and R. Alami. Situation assessment for human-robot interaction. In *20th IEEE International Symposium in Robot and Human Interactive Communication*, 2011.
35. M. Skubic, D. Perzanowski, S. Blisard, A. Schultz, W. Adams, M. Bugajska, and D. Brock. Spatial language for human-robot dialogs. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 34(2):154–167, 2004.
36. M. Tambe. Towards flexible teamwork. *JAIR*, 7:83–124, 1997.
37. S. Tellex. *Natural Language and Spatial Reasoning*. PhD thesis, MIT, 2010.
38. J. Trafton, N. Cassimatis, M. Bugajska, D. Brock, F. Mintz, and A. Schultz. Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics*, Part A:460–470, 2005.
39. B. Tversky, P. Lee, and S. Mainwaring. Why do speakers mix perspectives? *Spatial Cognition and Computation*, 1(4):399–412, 1999.