

A study of interaction between dialog and decision for human-robot collaborative task achievement

Aurélie Clodic[†], Rachid Alami[†], Vincent Montreuil^{†‡}
LAAS-CNRS,[†]
Université de Toulouse
7, Avenue du Colonel Roche,
31077 Toulouse, France
Université Paul Sabatier[‡]
118, route de Narbonne, 31062 Toulouse, France
aurelie.clodic@laas.fr

Shuyin Li, Britta Wrede, Agnes Swadzba
Applied Computer Science,
Faculty of Technology
Bielefeld University,
35594 Bielefeld, Germany
shuyinli@techfak.uni-bielefeld.de

Abstract—Human-robot collaboration requires both communicative and decision making skills of a robot. To enable flexible coordination and turn-taking between human users and a robot in joint tasks, the robot’s dialog and decision making mechanism have to be synchronized in a meaningful way. In this paper, we propose an integration framework to combine the dialog and the decision making processes. With this framework, we investigate various task negotiation situations for a social robot in a fetch-and-carry scenario. For the technical realization of the framework, the interface specification between the dialog and the decision making systems is also presented. Further, we discuss several challenging issues identified in our integration effort that should be addressed in the future.

I. INTRODUCTION

Service robots, like Jidowanki, Rackham or Biron (figure 1) need to be interactive. More specifically, they should be able to communicate their perception, system states, abilities etc. with human users. Not only do they have to provide this information to users on demand, but they also need to take initiative given an appropriate interaction context. These abilities are particularly important in collaboration tasks between a human user and a robot because they require frequent negotiation about execution status, modification of joint goals, task division and so on. The realization of sophisticated interactive capabilities presupposes a flexible dialog system that enables mixed-initiative interaction style and a powerful decision making mechanism that takes into account interaction in its planning process. We already realized a decisional framework for human-robot collaborative task achievement [2], [1], [7], [6], [3] and a grounding-based dialog system [14], [13]. Both systems work well individually and the challenge now is to integrate them into one single robot system to account for the complexity of collaborative tasks in Human Robot Interaction (HRI).

The integration of a decision making system and an interaction system for HRI poses new scientific questions

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

in comparison to traditional conversational planning applications in the field of Human Computer Interaction (HCI), e.g., [4], [16]. The major difference lies in the crucial role of environments. A mobile robot is situated “here and now” and is subject to environmental conditions. This characteristic means that the task planning in HRI can not be done purely based on pre-defined task rules, as popular in HCI. Instead, the planning process has to take into account unexpected environmental changes and supports instant plan modification. For example, a robot is supposed to fetch a cup of tea from a cafe to a user’s office, but detects the user on its way back to the office. The question now is: should the robot still bring the tea to the office, as planned originally, or ask the user to take the tea already in the corridor? Such issues are typical for HRI planning and requires constant observation of the environment, flexible planning mechanism and sophisticated interaction capabilities.

In this paper, we investigate the coordination between the dialog system and the robot supervision system, which is the central module of a robot and possess up-to-date information on the current environment and planning status. More specifically, we consider various cases of task negotiation with the focus of agreement building between a user and a robot during the collaboration. Further, we propose a framework that enables flexible coordination between the dialog and the supervision and present the interface specification between them. The considerations were made within the scenario Curious Robot (key experiment II of the COGNIRON project, <http://www.cogniron.org>).

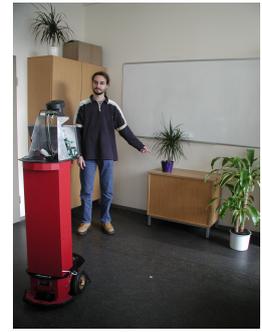
The paper is organized as follows: section II presents the integration framework for dialog and supervision. Section III discusses various planning-related interaction situations within the Curious Robot Scenario and how they can be handled using the framework. Then, section IV shows the interface specification between dialog and supervision that facilitates the technical realization of the framework. In section V, we discuss several challenging issues that were identified in our work and need to be addressed in the future. Finally, section VI summarizes the contributions of the paper.



(a) Jidowanki



(b) Rackham



(c) Biron

Fig. 1. Robots Jidowanki, Rackham and Biron in interaction context

II. AN INTEGRATION FRAMEWORK

Our framework is based on the joint activity theory that is often used to model collaborative tasks. In this section, we first briefly present this basic idea in Part II-A and then forward to the description of the general communication scheme between the dialog and the supervision in Part II-B. In Part II-C we discuss the possible communicative acts that are exchanged between the dialog and the supervision.

A. The basic idea

The chosen framework is the joint activity paradigm [8], [9], [10]. The central construct of this paradigm is *Joint Persistent Goal* (JPG). Agents form teams by adopting JPGs to achieve a collaborative task. JPG's hold if and only if all team members mutually believe:

- the goal is not yet achieved
- they want the goal to be achieved
- until the goal is known to be achieved, unachievable, or no longer relevant, they should persist in holding the goal

If a team member considers the goal to be achieved, unachievable, or no longer relevant, it will tell its teammates.

The teamwork concept is a widely-accepted metaphor for agent-agent interaction. In the current work, we use it to model human-robot interaction with the focus on interactive/collaborative task achievement by adding the following constraints:

- Human-robot interaction is oriented for task achievement,
- Human and robot share the space,
- Human and robot perceive each other,
- Human-robot Communication is based on Multi-modal dialog.

Note, unlike in some other applications, e.g., [11], [17], [18], we view HRI as unmediated, face-to-face interaction. A robot is considered as an autonomous agent that is able to manage interaction by itself.

The framework is designed to facilitate the realization of socially acceptable behaviors for a robot. More specifically, it should account for the following [5]:

- **Inter-predictability:** i.e. the need for the robot to behave such as its intentions are legible, and to predict what could be the human next action. The elaboration “recipes” is one step toward this goal.
- **Common ground:** Pertinent mutual knowledge and beliefs, as well as assumption about others’ abilities.
- **Directability:** capacity for the human to modify and for the robot to propose modification of recipes as conditions change; responsiveness of the robot to the human. This can be a response to an explicit request or to a situation that has been autonomously detected by the robot (e.g., human commitment monitoring).

B. Communication schemes

The supervision system supports two communication schemes that differ in terms of initiator of a communicative act: either the user or the robot (figures 2(a) and 2(b)). These schemes represent a possible sequence of events. These communication schemes can be compared to Sidner’s artificial discourse language [15] and Kumar’s protocols for joint actions [12]. What we added here is that we also consider possible misunderstandings (as well as their modelings) and adapted procedures to deal with.

Concerning the task division between dialog and supervision, the dialog is responsible for disambiguation of communicative acts on the language level. For example, a user issues an unclear utterance so that the dialog needs to initiate clarification questions such as “I beg your pardon?” However, it is sometimes not so easy to decide whether these language problems also have consequences for the planning process. Therefore, the supervision should be “warned” every time when the dialog receives and processes something even if the supervision is not directly involved.

During an interaction, the dialog system negotiates with the user till she provides a clear reply which the dialog can associate with a set of pre-defined supervision requests. This means that the supervision will always get a clear reply of the user. This process is performed along the task hierarchy that is produced by the planner or encoded as procedural knowledge in the supervisor (Open-PRS provides a context-dependent goal decomposition). Supervision and

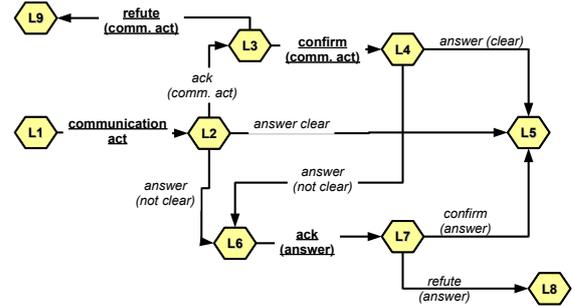
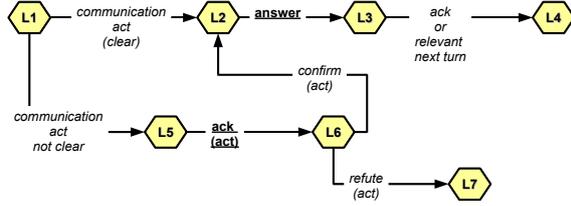


Fig. 2. Communication schemes : 2(a) when the human initiates the communication, 2(b) when the robot initiates the communication (given the answer : $val1 = val2$ or $val1 \neq val2$). **under line and bold** are robot acts and *italic* represents human's act

dialog interact about the SuperTask and joint subtasks, but not on sub-tasks that involve only the robot.

C. Possible communicative acts

In HRI, communication can be viewed as exchanges of communicative acts. We have defined a set of communicative acts that the human could do at every moment: make a request, suspend/resume a particular task, modify the plan etc. At any time, both the user and the robot can propose the following task-based communication acts:

- ASK_TASK: proposing a task,
- PROPOSE_PLAN: proposing a plan (recipe) for a certain task,
- PROPOSE_MODIFY_PLAN: proposing a modification of the current plan for a certain task,
- GIVE_UP: gives up a task (e.g., because the task becomes impossible). For the robot this is a way to announce that it is unable to achieve the task.
- CANCEL: cancellation of a task (voluntary give-up),
- TASK_DONE: announces that the task has been done,
- REALIZE_TASK: announces that the task performance will start.

Additionally, we have identified three interaction patterns that are typical for planning in our scenario:

- Direct vs. indirect replies: Depending on the situation an answer may follow directly a request or not. In such a case, there is an acknowledgment and the answer comes later.
- Explicit vs. implicit communication: Depending on the context, the communication acts may be implicit. For example, in some situations, it makes sense for the robot to proceed without asking the user or explicitly proposing a plan/recipe because the robot's intention is considered to be clear to the user. This is a case of implicit communication. Another example is If a person retracts his hand while the robot is passing him an object, then such behaviors of the user are viewed as an implicit communicative act CANCEL.
- "public" vs. "private" plans: The supervision system consider plans as constructing of two parts: a "public" and a "private" part. The public part has to be shared

with the robot's participant of the task (i.e., the user) and every modification of it should be negotiated. The private part does not need to be communicated with the user, e.g., robot changes its motion algorithm to adapt to the current environmental conditions.

III. SCENARIO AND VARIATIONS

This section discusses variations of a fetch-and-carry task in the Curious Robot Scenario. The goal of this discussion is to demonstrate how different interaction situations can be modeled with the framework proposed above.

The basic scenario is the following. Robot Jidowanki (called Jido below) is in its waiting zone. User Luis comes and asks Jido: "robot, bring me a coke". Jido accepts. It goes to the cafe, finds the cola can, and brings back the cola to Luis' office. Jido passes the cola to Luis.

We defined a first sketch of the interaction protocol and data exchanges between Dialog and supervision for human-robot interactions for "joint tasks". The planner intervenes as a "resource" that is invoked by the supervision, when necessary, to assess the feasibility of a task in a given context. Below, the cooperation between the dialog, supervision and planning system are illustrated using UML sequence diagram.

A. Starting and Ending the interaction

At the beginning, Luis asks the robot to bring him a cup of coffee. This utterance is interpreted by the dialog system as a communicative act ASK_TASK with parameters $object=coffee$ and $person=Luis$. After this act is forwarded to the supervision system, it consults the planner to find out whether a plan for the task is available. Based on the reply of the planner, the dialog system informs the user as to whether the robot accepts or has to reject the task. This process is illustrated in figure 3.

Both the user and the robot are able to initiate the end of an interaction. In either case, a common ground concerning the task execution status has to be established between them. For this purpose the interaction partner who did not initiate the end of interaction has to explicitly reply to signal their agreement or disagreement (see figure 4).

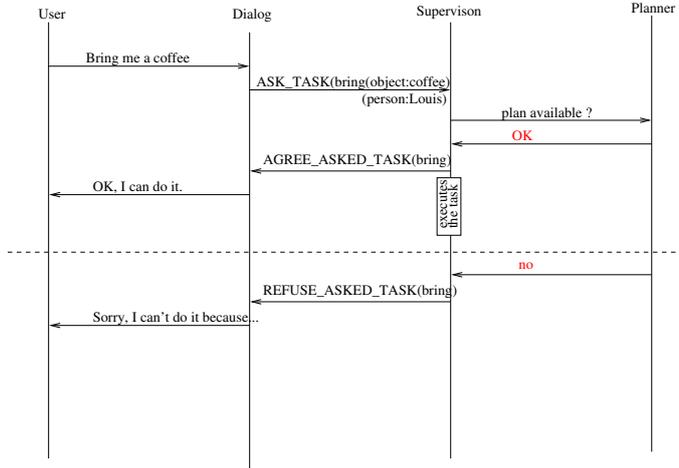


Fig. 3. Interaction start: User initiates a task and the robot accepts or refuses it depending on the availability of an plan

B. Robot-initiated plan modification

Consider the following variations of the basic scenario:

[Example 1] Jido meets Luis at the cafe. Before it takes the can, the robot proposes Luis to take the can himself. Two possibilities: Luis agrees or refuses.

[Example 2] Jido meets Luis in the corridor while heading to Luis Office after it has taken the can. Jido proposes to Luis to take the can immediately. Luis can agree and takes the cola can, or refuse the robot's proposal.

[Example 3] Jido goes to the cafe to find out that there is no cola anymore, but cola is still available in the restaurant nearby (Luis did not specify where to get the cola).

In example 1 and 2 the robot modifies the public part of its plan and in example 3 the private part. In all the three cases, a new plan has to be constructed because of unexpected environmental changes. During the processing, the supervision system perceives the environmental changes and the planner makes the decision that a new plan needs to be constructed. Upon receiving the information on the modified plan, the dialog system takes the initiative to proposes it to the user.

Figure 5 illustrates how the example 2 is handled in our framework. The unexpected presence of the user in the corridor triggers an *irrelevance condition* so that the supervision system asks the planner whether a new plan is available. Irrelevance does not mean that the current plan is impossible, but that a better one may be selected given the new task context. If a new plan is available, an act `MODIFY_PLAN` is sent to the dialog system and the robot now waits for the reply of the user. If he agrees, then the new plan will be employed; if not, the old plan will be resumed.

C. User-initiated plan modification

Now, we analyze what happens if the user proposes to modify the robot's plan or specifies a part of it.

[Example 1] U: "Bring me a coffee. Put it on my desk"

[Example 2] U: "Bring me a coke to my office. Oh, but put it on my desk (not on my chair)"

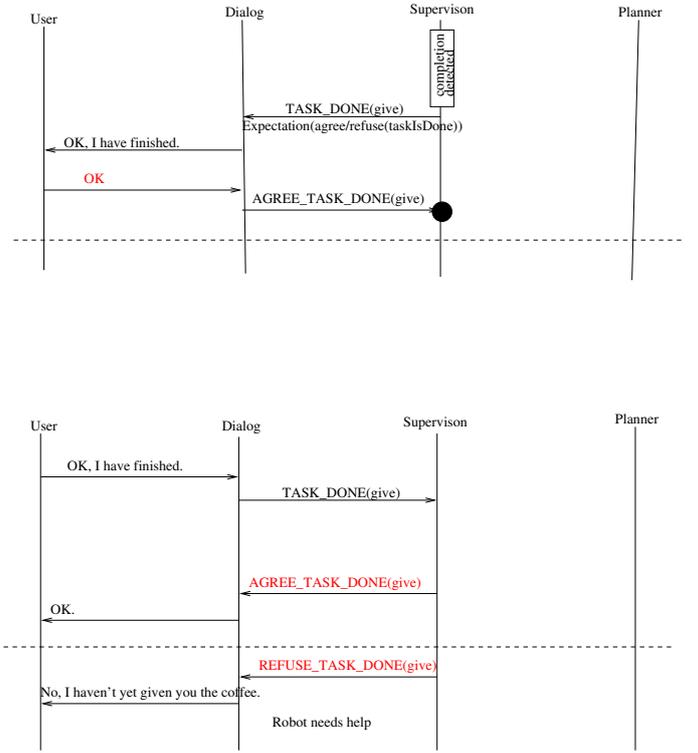


Fig. 4. Interaction End: the first and second diagram illustrate the processing flow in case that robot and the user initiates interaction end, respectively

[Example 3] U: "Bring me a coffee. Give it to me."

[Example 4] Jido arrives at Luis office, Luis said "put it there please".

Here the user modifies the robot's plan by specifying a part of it. The dialog system would identify "bring me a coffee!" and "put it on my desk." in the example 1 as two plans rather than one plan plus a plan modification. Even though, the supervision system should have the possibility to access the plan definition of "bring" so that it can identify that "put" is a subtask of "bring" and the user is probably proposing a modification of the previous task.

Figure 6 illustrates how the example 4 is handled in our framework. A part of the robot's plan of delivering an object comprises passing the object to the user. According to this plan, the robot will ask the user to take the object, in this case a cup of coffee, after it arrives at his office. The idea case is that the user agrees, i.e., shows his commitment to the robot's proposal by saying "ok, go on" and stretching his hand to take the coffee. However, if the user rejects the robot's proposal and specifies a part of the plan instead, e.g., by saying "No, put it on my table", then the supervision system handles it by asking the planner to construct a new plan.

IV. INTERFACE SPECIFICATION

To realize the integration framework for the Curious Robot scenario, we specified the interface between the dialog system and the supervision system in XML.

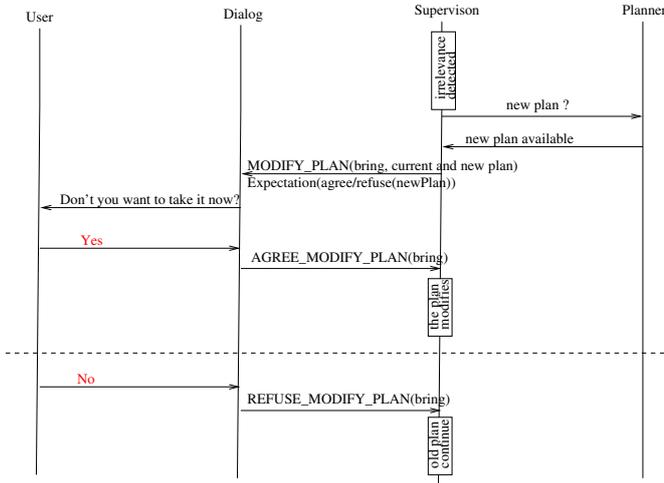


Fig. 5. Robot-initiated plan modification (for example 2): The robot encounters the user before reaching the place where it was supposed to give the can to him. The robot then proposes to give user the can in the current place

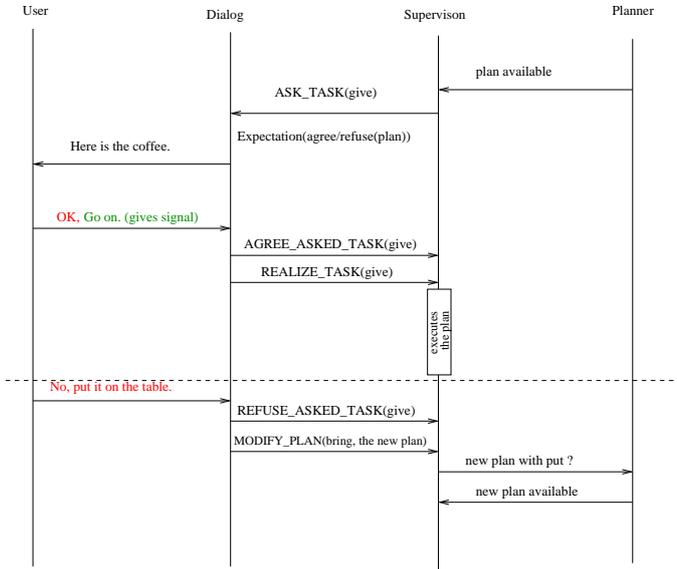


Fig. 6. User-initiated plan modification (for example 4): the robot hands a coffee to the user. Either the user agrees and acts as planned by the robot, or he rejects and proposes an alternative solution for the same joint task: put the coffee on the table.

A. From the supervision system to the dialog system

```
<MSG>
<sender>Supervision</sender>
<receiver>DLG</receiver>
<commAct state = "launch/stop">
  <act>the communicative acts as
  specified in section II
  </act>
  <property>agree or refuse or act</property>
  <info>
    <task id = "the ID of a task
    (specified by the user)
    or a subtask
    (specified by the planner)"
  >
    <person>...</person>
    <obj>...</obj>
```

```
...
</task>
<plan feature = "new or current">
  the name of the plan
</plan>
...
</info>
</commAct>
<expectation>
  <commAct> as defined above</commAct>
  ...
</expectation>
</MSG>
```

The attribute *state* of the tag *commAct* can have two values: either *launch* or *stop*. They indicate whether the supervision system is launching or stopping a communicative act. If the value of the tag *property* is *act*, it means that it is an act that should be processed.

B. From the dialog system to the supervision system

```
<MSG>
<sender>DLG</sender>
<receiver>Supervision</receiver>
<commAct state = "final/negotiating">
  <act>the communicative acts as
  specified in section II
  </act>
  <property>agree or refuse or act</property>
  <info>
    <task>
      <person>...</person>
      <obj>...</obj>
      <position>the concrete position or
      gestureExpected
    </position>
    ...
  </task>
  <plan feature = "new or current">
    the name of the plan
  </plan>
  ...
  </info>
</commAct>
</MSG>
```

The attribute *state* of the tag *commAct* can have two values: either *final* or *negotiating*. They indicate whether the dialog system has already finished the negotiation of the current utterance or not, i.e., whether the current message is a final result of the language level analysis and it contains all the information specified by the user.

As to the tag *act*, if the dialog system can identify it, then it is set to this tag. The tag *position* contains the value *gestureExpected* if the dialog system expects an gesture based on key words in user utterances, e.g., this or that, but no gesture can be found in the gesture-database. Upon receiving this message, the supervision system will activate the gesture recognition system and store its results into the database to which the dialog system also has access. If the gesture recognition fails to deliver any results, the dialog system initiates clarification questions to ask the user to do the gesture again.

V. CHALLENGING ISSUES

The scenario variations discussed in the section III can be already taken into account with our framework and the

interface specification. However, we also identified several challenging cases that are still need to be studied in the future. This section provides an overview of these problems.

In case that the robot can not construct a plan for a task specified by the user, it is ideal if the robot could explain the reason for it. Such behaviors would help users to understand the system and increase its usability. However, most standard planning systems today are self-contained and can not provide such information.

Sometimes users specify a task and complement it with more information that is true only in the future, e.g., "Bring me a coffee. I will be in my office". Many planning systems can not handle such situations. We are working on a planning system named HATP that is able to synchronize plans with expected events by performing temporal planning.

In general, it is challenging for the dialog system to identify relationships between tasks. For example, if the user says "bring me a coke ah... and bring me a mars", then the dialog system can interpret the utterance as either two tasks with two individual goals, or two tasks with one single goal or a task and a subtask. The correct interpretation of such utterances requires detailed representation of the task planner, so that the dialog system can view the world in the same way as the planner. However, this would result in double representation of tasks in the robot system and has consequence for the efficiency of the system.

Robots should be able to communicate their abilities with users. It is quite simple for a robot to answer questions like "What can you do?" because it only concerns the general capabilities of a robot and it can be realized even with a standard reply. However, if the robot is asked "What can you do *now*?" or "What can you do *if I do x*?", it is a totally different case. To answer such conditional questions the system must be aware of its current state and the state of the environment. Further, it requires a unified representation of this information that both the planner and the dialog system have access to.

Last but not least, the understanding of implicit commands such as "Oh, it's cold here! No wonder, the window is open" is highly challenging. To realize it, the main question is which module should be responsible for inducing the true meaning of such commands. If it should be the dialog system, then it must possess sophisticated reasoning capability to interpret the example command as ASK_TASK(open(window)). However, such reasoning capabilities are most of the time realized by task planners and the dialog system often only performs language-level analysis. If the planner should induce the meaning, then it is the question how should the dialog system represent such utterances.

VI. CONCLUSION

In this paper we proposed an integration framework for the dialog, supervision and planning system of a robot. The goal of this integration effort was to realize sophisticated interactive and planning capabilities for mobile robots that act in real environments and, therefore, have to cope with various unexpected situations. The framework is based on

the joint activity theory of Cohen and is adapted to special needs of HRI. We analyzed various interaction situations and demonstrated how they can be handled using this framework. Further, we also proposed the interface specification between the dialog and the supervision system, which contribute to the technical realization of the integration in a concrete way. The framework is currently being implemented for the robot Jidowanki and Biron and, in the future, we will address the challenging issues identified in our current integration effort.

REFERENCES

- [1] R. Alami, A. Clodic, V. Montreuil, A. Sisbot, and R. Chatila. Towards human-aware cognitive robotics. In *Cogrob 2006, The 5th International Cognitive Robotics Workshop (The AAAI-06 Workshop on Cognitive Robotics)*, Boston, USA, 2006.
- [2] R. Alami, A. Clodic, V. Montreuil, A. Sisbot, and Chatila R. Toward human-aware robot task planning. In *AAAI Spring Symposium "To boldly go where no human-robot team has gone before"*, 2006.
- [3] R. Alami, A. Clodic, V. Montreuil, E. A. Sisbot, and R. Chatila. Task planning for human-robot interaction. *The Smart Objects and Ambient Intelligence Conference, sOc-EuSAI 05, Grenoble, France*, 2005.
- [4] J. Allen, D. Byron, M. Dzikovska, G. Ferguson, L. Galescu, and A. Stent. Towards conversational human-computer interaction. *AI Magazine*, 22(4), 2001.
- [5] Jeffrey Bradshaw, Paul J. Feltovich, Hyuckchul Jung, Shrinivas Kulkarni, James Allen, Larry Bunch, Nathanael Chambers, Lucian Galescu, Renia Jeffers, Mathew Johnson, Martien Sierhuis, William Taysom, Andrzej Uszok, and Ron Van Hoof. Policy-based coordination in joint human-agent activity. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 2004.
- [6] A. Clodic, S. Fleury, R. Alami, R. Chatila, G. Bailly, L. Brthes, M. Cottret, P. Dans, X. Dollat, Elisei F., Ferran I., M. Herrb, G. Infantes, Lemaire C., F. Lerasle, J. Manhes, P. Marcoul, P. Menezes, and V. Montreuil. Rackham: An interactive robot-guide. In *IEEE International Workshop on Robots and Human Interactive Communication (ROMAN)*, Hatfield, UK, 2006.
- [7] A. Clodic, V. Montreuil, R. Alami, and R. Chatila. A decisional framework for autonomous robots interacting with humans. *IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN)*, 2005.
- [8] Philip R. Cohen and Hector J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42(2-3):213–361, 1990.
- [9] Philip R. Cohen and Hector J. Levesque. Teamwork. *Nous*, 25(4):487–512, 1991.
- [10] Philip R. Cohen, Hector J. Levesque, and Ira Smith. On team formation. *Contemporary Action Theory*, 1998.
- [11] Terrence W Fong, Clayton Kunz, Laura Hiatt, and Magda Bugajska. The human-robot interaction operating system. In *2006 Human-Robot Interaction Conference*. ACM, 2006.
- [12] Sanjeev Kumar, Marcus J. Huber, and Philip R. Cohen. Representing and executing protocols as joint actions. In *Proceedings of the First International Joint Conference on Autonomous Agents & Multi-Agent Systems (AAMAS)*, 2002.
- [13] Shuyin Li and Britta Wrede. Why and how to model multi-modal interaction for a mobile robot companion. In *Proc. AAAI Spring Symposium on Interaction Challenges for Intelligent Assistants*, Stanford, 2007. to appear.
- [14] Shuyin Li, Britta Wrede, and Gerhard Sagerer. A computational model of multi-modal grounding. In *Proc. ACL SIGdial workshop on discourse and dialog, in conjunction with COLING/ACL 2006*, pages 153–160. ACL Press, 2006.
- [15] Candice L. Sidner. An artificial discourse language for collaborative negotiation. In Barbara Hayes-Roth and Richard Korf, editors, *Proceedings of the Twelfth National Conference on Artificial Intelligence*, pages 814–819, Menlo Park, California, 1994. AAAI Press.
- [16] R. Smith and D. R. Hipp. *Spoken natural language dialogue systems: A practical approach*. Oxford University Press, 1994.
- [17] Milind Tambe. Agent architectures for flexible, practical teamwork. *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 1997.
- [18] Milind Tambe. Towards flexible teamwork. *Journal of Artificial Intelligence Research*, 7:83–124, 1997.