Symmetrical joint action in human-robot dialogue

Mary Ellen Foster Tomas By Markus Rickert Alois Knoll
Robotics and Embedded Systems Group
Department of Informatics, Technical University of Munich
Boltzmannstraße 3, 85748 Garching, Germany
{foster,by,rickert,knoll}@in.tum.de

***Abstract*— We describe a model for task-driven, natural-language-based human-robot interaction where the roles of the two participants are fully interchangeable, both at the conversational level and at the task level. This model provides a general treatment of joint-action dialogue, and also allows models based on the analysis of human-human joint-action dialogues easily to be implemented and evaluated.**

I. Dialogue and joint action

According to Clark [1], natural-language dialogue has its origins in joint activity, and serves to coordinate it. Whenever agents work together, they must agree on the goals and subgoals, the means to address each goal, and which agent should perform which tasks; all of these agreements are reached through dialogue. Dialogue incorporates joint activity on two levels: dialogue in service of the joint activity, and the joint activity of managing the dialogue itself.

When a human and a robot cooperate on a joint task, the most intuitive interface is one that provides the same coordination abilities as human-human dialogue. In most existing human-robot dialogue systems, the participants have very different roles and capabilities, which limits the range of potential joint activities. In this paper, we propose that fully symmetrical human-robot cooperation is a promising line of research as well, as it is able to provide a direct mapping from a wider range of human joint activity to the human-robot case, and thus allows interaction models based on human-human interaction to be implemented and tested.

A. Human-robot joint activity

Traditionally, a robot has been seen as a tool for accomplishing specific tasks. When this view is taken, the obvious human-robot interface involves the human giving commands and the robot carrying out the requested actions, with no support for more sophisticated communication. For many applications, such an interface is sufficient; however, as the tasks to be performed by robots grow more complex, natural-language dialogue provides a more natural and flexible interface that does not suffer from problems such as limited screen size.

In most systems that permit natural-language human-robot interaction, the roles of the human and robot agents are hard-coded and distinct, and there is a limited amount of collaboration beyond that which is needed to manage the dialogue. The WITAS robot helicopter [2], for example, is an autonomous vehicle operating at a distance from the user, and the majority of communication consists of the user giving commands to the vehicle. The main form of joint activity in this sort of system occurs when the robot must ask the user for help in dealing with a situation—e.g., confirming that a recognised object is of the intended type. The Human-Robot Interaction Operating System (HRI/OS) [3] supports peer-to-peer interaction for human-robot teams cooperating on operational tasks such as seam welding, communicating through dialogue. More communication and direct collaboration is possible among these agents than in WITAS, but the agents still play different, complementary roles. A somewhat different system is Mel the robotic penguin host [4], who guides the user as they explore a lab. The main goal for Mel is to create social engagement in a conversation; collaboration takes place when the robot instructs the user on using the equipment in the lab, but the roles are still distinct and fixed.

Other human-robot systems permit more symmetrical collaboration between robots and humans, with more flexible role assignment, but generally provide less support for natural-language dialogue. For example, Leonardo [5] is a fully-embodied humanoid robot with social skills that allow it to learn and collaborate effectively in human settings. Leonardo is able to respond to requests from a human, to learn both new procedures and the names of objects, and to execute learned procedures in collaboration with the user. The emphasis is on the robot learning and executing action sequences; Leonardo cannot produce spoken output, and communicates through actions, body language and facial expressions. The eventual goal of the Ripley system [6] is also to support collaborative human-robot interaction; however, most of the work so far on that system has concentrated on the development of a mental model to ground symbolic processing in sensor data.

B. Joint activity in human-human dialogue

Clark [1] defines five dimensions along which joint-activity types may vary: scriptedness, formality, verbalness, cooperativeness, and governance (i.e., the eqality or inequality of the participants' roles). The human-robot systems described in the preceding section mainly cover only a part of this space; the systems where a human directs an autonomous vehicle have unequal role assignments and use mainly verbal communication, for example, while the Leonardo system has more equal roles but is unable to produce verbal output.

If a human-robot dialogue system supports a wider range of human-human joint activities, this permits interaction strategies based on human behaviour to be modelled; for exam-

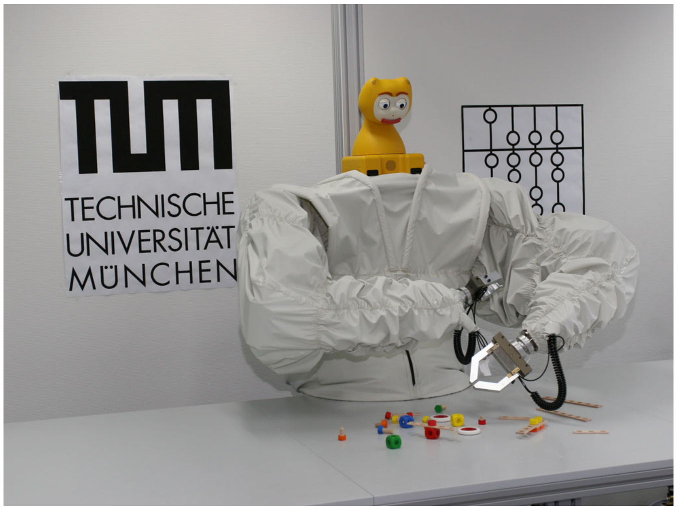

Fig. 1. The JAST human-robot dialogue system

ple, the system may follow human patterns for determining roles, allocating subtasks, or dealing with failed actions or ambiguous instructions. The resulting system can then be evaluated to determine the impact that such human-derived strategies have on the dialogue, using both objective task-success measures and subjective measures of satisfaction and engagement. To support this range of human-human joint activities requires a more flexible view of role assigment, along with the ability to coordinate activity at all levels, using natural language and other communication channels. In the next section, we describe a human-robot dialogue system that takes this approach.

II. The JAST human-robot dialogue system

The overall goal of the JAST project ("**J**oint **A**ction **S**cience and **T**echnology"; http://www.euprojects-jast.net/) is to investigate the cognitive and communicative aspects of jointly-acting agents, both human and artificial. The human-robot dialogue system being built as part of the project is designed as a platform to integrate the project's empirical findings on cognition and dialogue with its work on autonomous robots.

In the JAST dialogue system [7], a robot and a human user work jointly to assemble a wooden construction toy (Figure 1), coordinating their actions through speech, gestures, and facial motions. The interaction incorporates joint action at all levels: the user and the human jointly determine the goals to pursue and the procedures to be used to pursue them, and also work together to carry out the selected plans. The roles of the two agents in the JAST system are, in principle, completely interchangeable: either may make, accept, or reject proposals, and either—or both—may also carry out domain actions such as manipulating and assembling components.

Dialogue management in JAST [8] is based on Blaylock and Allen's *collaborative problem-solving* (CPS) model of dialogue [9], which represents dialogue as an interleaved, collaborative process of selecting the goals to pursue, determining procedures for pursuing those goals, and executing the selected procedures. The central process in the CPS model is the selection of values to fill roles; for example, agents must decide on the top-level objective to pursue, the recipe for pursuing it, and the specific objects in the world that should be used in the recipe. This model is particularly suited to the sort of collaborative task-driven dialogues supported in JAST.

We intend to exploit the symmetry of the agents' capabilities to implement problem-solving and interaction strategies based on recordings of humans engaged in similar joint-action construction tasks; partners in the JAST project are currently recording and analysing such interactions. Data from these recordings will provide information on natural strategies for role assignment and negotiation, grounding, and confirmation in joint-action dialogues. We will assess the success of the implemented strategies through human evaluations.

III. Conclusion

We have outlined the main features of human-human joint-action dialogue, and proposed that an intuitive interface for a human-robot system is one that supports this sort of human-like collaboration between two agents with similar abilities and equal initiative in the dialogue. Giving the robot agent similar abilities to the human allows interaction patterns and strategies from human-human dialogue to be straightforwardly modelled, implemented and compared with more hard-coded strategies. We have also described the JAST human-robot dialogue system, which aims to support this type of symmetrical cooperation on a joint-action construction task, using a dialogue manager based on collaborative problem solving.

References

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

[2] O. Lemon, A. Gruenstein, and S. Peters, "Collaborative activities and multi-tasking in dialogue systems," *Traitement Automatique des Langues (TAL)*, vol. 43, no. 2, pp. 131–154, 2002.

[3] T. Fong, C. Kunz, L. M. Hiatt, and M. Bugajska, "The human-robot interaction operating system," in *Proceedings, 1st Annual Conference on Human-Robot Interaction (HRI 2006)*, 2006.

[4] C. L. Sidner and M. Dzikovska, "A first experiment in engagement for human-robot interaction in hosting systems," in *Advances in Natural Multimodal Dialogue Systems*, N. O. Bernsen, L. Dybkjær, and J. van Kuppevelt, Eds. Springer, 2005.

[5] C. Breazeal, A. Brooks, J. Gray, G. Hoffman, C. Kidd, H. Lee, J. Lieberman, A. Lockerd, and D. Chilongo, "Tutelage and collaboration for humanoid robots," *International Journal of Humanoid Robotics*, vol. 1, no. 2, pp. 315–348, 2004.

[6] D. Roy, K.-Y. Hsiao, and N. Mavridis, "Mental imagery for a conversational robot," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 34, no. 3, pp. 1374–1383, 2004.

[7] M. E. Foster, T. By, M. Rickert, and A. Knoll, "Human-robot dialogue for joint-action construction tasks," in *Proceedings, 8th International Conference on Multimodal Interfaces (ICMI 2006)*, 2006, to appear.

[8] M. E. Foster, "Dialogue management for cooperative task-driven human-robot dialogue," in *Proceedings, Brandial 2006 poster session*, 2006, to appear.

[9] N. Blaylock and J. Allen, "A collaborative problem-solving model of dialogue," in *Proceedings, 6th SIGdial Workshop on Discourse and Dialogue*, L. Dybkjær and W. Minker, Eds., 2005, pp. 200–211.