Building a Bridge with a Robot: A System for Collaborative On-table Task Execution

Ruth Schulz
Machine Learning and Robotics Lab, University of Stuttgart, Germany
Ruth.Schulz@ipvs.uni-stuttgart.de

Philipp Kratzer
Machine Learning and Robotics Lab, University of Stuttgart, Germany
st101215@stud.uni-stuttgart.de.

Marc Toussaint
Machine Learning and Robotics Lab, University of Stuttgart, Germany
Marc.Toussaint@ipvs.uni-stuttgart.de

ABSTRACT
It is possible to build highly accurate and mobile robots able to perceive their environment. Such robots can be used to assist humans in everyday tasks to reduce their workload. As a consequence, communication and interaction between humans and robots is becoming more important. In this paper we present a system for human-robot collaboration for on-table tasks. Due to its extensible design our system serves us as a base for further investigations in human-robot collaboration. We used the system to implement five action-selection strategies for the robot: proactive, autonomous, reactive, human-requested and human-commands. We conducted a pilot study to compare the interaction modes during a task in which the human and the robot build a bridge using blocks. The results of the pilot study indicate that for the simple bridge-building task, people prefer to interact with a robot using the proactive action-selection strategy. The completion of the pilot study indicates that the system is useful for human-robot collaboration studies. Several limitations have been identified that will be addressed in future developments.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous;

Author Keywords
Robot; Tabletop; Human-Robot Interaction

INTRODUCTION
While industrial robots of the past were designed to perform their tasks in isolation from humans to prevent accidents, collaborative robots are now being designed to interact directly with humans in a wide variety of situations. Robots can be highly accurate, mobile, and perceive many details of their environments, but if robots are to assist humans directly, communication and interaction are important factors. It is still an open question how people want to interact with robots when they are collaborating together on a task. How and when should a robot act? What role should a robot play? When should a robot watch or ask questions rather than act?

HAI 2017, October 17–20, 2017, Bielefeld, Germany..

ACM ISBN 978-1-4503-5113-3/17/10 ...$15.00.
http://dx.doi.org/10.1145/3125739.3132606

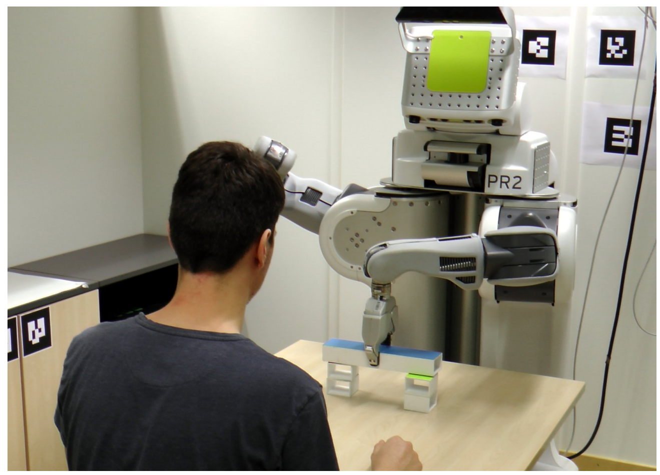

Figure 1: Human and robot performing an on-table task.

For collaborative armed robots, on-table tasks is one domain in which they can be used (see Figure 1). These robots need to be able to interact and perceive the table, objects on the table, the human, and changes to the environment. Ideally, they would be able to work out what task is being completed, and how the human wants them to help. This involves understanding requests from the human and working out when they should act, watch, and ask questions.

Human-robot interaction is a growing field with a large variety of research included under the theme. Research into designing robots for effective human-robot interaction has taken inspiration from human interactions in teams [6, 10]. The automated planning field has mainly advanced in simulated tasks, but it is now possible to use some of the techniques developed in this field for robot action-selection in real human-robot interactions [8]. Other studies have investigated how robots should choose when to interact [9, 1], in particular whether a robot should wait to be instructed or should be proactive in selecting actions.

In developing robot systems for studying human-robot interactions, a key consideration is the design of the task to be collaboratively completed by the human and the robot. The ideal task is sufficiently complex to allow interesting collaborative actions to be investigated, while remaining simple enough that a useful analysis of the interactions can be performed. Past studies have often involved simple tasks that do not actually require collaborative actions, for example, tidying up. This can result in interactions that involve two agents acting in the same workspace, but not actively collaborating on a task.

The goal of this work is to develop a system for a collaborative armed robot to interact with objects on a table together with a human, to allow testing of algorithms and interaction modes on a simplified real world task. The table-top block world allows tasks to be designed for different levels of interaction: from independent action, to sequential actions performed by multiple agents, to collaborative actions requiring joint action by multiple agents.

This paper presents a system for human-robot collaboration for on-table tasks. The environment is a table-top block world for humans and robots to interact with each other. The system allows the efficiency of the algorithms to be tested as well as how humans want to interact with robots. The system can be extended with different tasks and robot action-selection techniques. In this paper, we present a pilot study using the system, comparing different interaction modes (proactive, autonomous, reactive, human-requested, and human-commands) for a task in which the human and robot build a bridge together using blocks.

The study demonstrates that the system can be used for table-top block world interactions between humans and robots. The results indicate that people prefer a proactive robot, as in previous research. We will continue to develop the system to conduct further investigations in collaborative on-table tasks.

The paper continues with a background of the relevant literature, then presents a description of the robot system, the experiment and results, then conclusions and future work.

RELATED WORK

A variety of challenges exist for human-robot collaborative tasks, including communication, joint action, and human-aware execution. One drive behind human-robot interaction research is to develop a robot that is able to interact with humans efficiently for successful completion of collaborative tasks. Such robots require geometric reasoning, situation assessment, knowledge models, dialog, task planning, and task execution [4]. Current challenges for designing robots for human robot interaction include bi-directional intent recognition, how to determine which role the robot should take in an interaction, how to choose an action with appropriate tradeoff between individual preferences and task completion, and how the robot can self-evaluate its own performance [3]. We focus on developing a system that can be used to test techniques addressing these challenges. The pilot study presented in particular addresses action selection based on the role that the robot has in the interaction.

Several studies have tested how humans want to interact with robots when performing collaborative tasks and how to develop robots with the abilities required to perform these tasks. Human intention has been used as a cue for proactive robot task selection [9]. Gaze cues have been used to support joint attention to aid in task completion, as well as tracking actions and behaviors to enable questions and feedback to improve task completion [5]. Another potential source of information for action selection is the mental state of the human partner with respect to goals, plans, and actions [2]. Knowledge of the human partner's mental state can inform action choice to allow the human to better understand the robot's capabilities [7]. While our system could be extended to include a variety of cues indicating human intention, the current implementation focuses on identification of the current state of the task with complete knowledge of the actions that need to be executed for task completion.

More is currently know about human teamwork practices, and how humans best interact with other humans, than how humans interact with robots. Some studies have used ideas from effective human teamwork practices to improve training of robots [6] and to develop strategies that aim to reduce human idle time [10]. Automated planning techniques have been integrated with a robot bartender choosing which action to undertake [8]. Three different policies for robot initiative were tested in a user study where the robots performed collaborative on-table tasks with a human, with the proactive policy found to be most efficient [1]. The interaction modes investigated in this pilot study include modes similar to those used in [1].

SYSTEM

In order to investigate human-robot collaboration in on-table tasks we developed a system that allows perception and manipulation of objects. We focus on the situation in which the human and the robot sit on opposite sides of the table (see Figure 1). The design of the system allows new tasks and action selection policies to be easily added.

Platform

We use the PR2 robotics and research platform. To perceive objects on the table we use the robot's stereo camera that is attached to the head. Furthermore, the robot has two 7-degree of freedom arms with attached grippers used here for on-table pick and place actions.

State and Actions

The state, S, of the system is described by the state of the table, S_{table}, and the agent states, S_{human} and S_{robot}. S_{table} consists of the position and orientation of the objects that are located on the table. Objects located in the hand/gripper of the agent are described by the agent states S_{robot} and S_{human}. The state S_{table} can be influenced by both agents while S_{robot} and S_{human} only can be influenced by the particular agent. Both agents can perform actions to manipulate objects on the table or influence actions from the other agent. Possible actions are pick an object, place an object, wait, and communicate. In our system, communication is only done by the human in form of commands. Both agents follow an action selection policy $\pi(S) = a$ that selects actions based on the current state and if a command C is considered based on the current state and command $\pi(S,C) = a$.

Implementation

The robot obtains point clouds of tabletop objects by using the depth values of the stereo camera. We use a discretization step to obtain discretized objects with a predefined size and color. This allows a better assignment of detected objects to real world objects. If an object is placed on top of another object, only one point cloud is detected. As a consequence, we need to keep track of the objects in our scene and add *below*-relations to an object in the case that another object is placed on top. The system is implemented in a way that first the perception step is performed. Afterwards the robot can select an action based on the information from the obtained discretized step.

EXPERIMENT

An important part of collaborative task execution is when an action should be selected and who should take initiative during the action selection process. We identified five main interaction modes and conducted a pilot study to compare these modes on a simple building blocks task. The pilot study also serves to identify current limitations of our system to be addressed in future developments.

Task

We implemented a task in which the robot and human should build a bridge in a collaborative manner (see Figure 1). The objective is to place five blocks in a predefined way to reach a target state. The target state is known to both agents. Due to the range of the robot not all of the objects can be picked by the robot. In such situations the human has to pick the object. Accordingly we assume that not all objects can be easily reached by the human. As a consequence, we have three areas of the table: a human-only, a robot-only and a common area.

Interaction modes

In this section we describe the five main interaction modes investigated in the study.

- **Autonomous** The robot in autonomous mode obtains the actions that are needed to reach the next subtask state. If one of the actions can be performed by the robot, the robot autonomously performs the action. If no action is possible, the robot will wait until an action becomes ready or the state changes.
- **Proactive** The robot in proactive mode is similar to the robot in autonomous mode. However, if none of the actions needed to reach the next subtask state can be performed by the robot, it will start doing an action that needs to be done later on. For example, the robot could grasp an object that needs to be placed on top of an second object that is not placed at that time.
- **Reactive** In reactive mode, the robot observes the state and monitors the human's actions. If the next subtask state is not reached in a predefined time window, the robot will assume difficulties in performing the action. As a consequence, the robot will perform an action leading to the next subtask, if possible. This mode is similar to the *reactive* mode by Baraglia et al. [1].
- **Human-requested** The robot observes the state. If the human requests help, the robot will check which actions reach the next subtask and will perform one of them. To simplify the implementation the researcher pressed a button when the human said the command "Robot, can you help?". This mode is similar to the *human-requested* mode by Baraglia et al [1].
- **Human-commands** The human can ask the robot to perform a certain action. If the action is possible, the robot will perform the action. To simplify the implementation the researcher pressed a button according to the command the participant said. For example, a possible command is "Robot, can you place the red block".

Apparatus

The participant and the robot sit on opposite sides of a table (see Figure 1). The table has a size of 80cm×80cm and a height of 70cm. Four small blocks of different colors of the size 6cm×6cm×4cm and a large block of the size 6cm×27cm×4cm were placed on the table. A camera behind the participant recorded the study. A smaller table with the task description and the questionnaire was placed near to the participant.

Procedure

After welcoming our participants, they filled in a consent form and a demographics questionnaire. The goal of the task was explained and the participant performed the task once without help of the robot. Each participant then performed the task with the robot five times, once for each interaction mode. The order of the interaction modes was randomized. The blocks were initially placed so that the human had one block that needed to be placed on top of a block placed by the robot and vice versa. To decrease learning effects we changed the initial position of blocks without violating that constraint. Before starting a mode we explained what the participant needed to do for the robot to assist. For the autonomous, proactive, and reactive modes, the participant was informed that the robot would determine when it should act, and that no further communication with the robot was needed. for the human-requested mode, the participant was instructed that they should say "Robot, can you help?" when they wanted the robot to do something. For the human-commands mode, the participant was instructed that they should ask the robot to perform a specific action, for example "Robot, can you place the red block". After each interaction mode the participant was asked to answer the following statements on a 7-point Likert scale:

- **S1**: The robot and I worked efficiently together
- **S2**: The collaboration with the robot felt natural
- **S3**: I was relaxed during the task execution
- **S4**: The robot was intelligent
- **S5**: I was surprised by what the robot was doing

	Proactive	Autonomous	Requested	Commands	Reactive
S1	5.71 (0.76)	5.57 (0.79)	4.57 (1.90)	5.29 (0.76)	4.14 (1.21)
S2	4.57 (1.27)	4.43 (1.27)	3.86 (1.57)	4.43 (0.98)	3.71 (0.76)
S3	5.57 (0.53)	5.57 (0.53)	5.29 (0.76)	5.43 (0.79)	5.43 (0.53)
S4	4.71 (0.95)	4.43 (0.79)	4.57 (0.98)	5.00 (1.15)	4.14 (0.90)
S5	1.71 (1.98)	2.57 (2.30)	2.00 (2.16)	2.00 (2.00)	1.71 (1.38)

Table 1: Rating of the collaboration with the robot regarding efficiency (S1), naturalness (S2), relaxation (S3), intelligence (S4), and surprise (S5) reported by the participants (From 0-*strongly disagree* to 6-*strongly agree*). Mean(SD)

To gain qualitative feedback about the system, the participant was asked for general comments and improvements for the system after the five tasks were completed.

Results

We conducted a lab study with 7 participants (1 female) aged between 21 and 51 ($M = 32.0$, $SD = 10.0$). The participants' responses to the statements regarding efficiency (S1) and naturalness (S2) of the collaboration, relaxation during the task (S3), intelligence of the robot (S4), and surprise about what the robot was doing (S5) are summarized in Table 1.

The reactive mode, in which the robot waits until a set time has elapsed, is perceived as less efficient ($M = 4.14$, $SD = 1.21$) and natural ($M = 3.71$, $SD = 0.76$) than the other modes. This coincides with the results by Baraglia et al [1].

The results show a trend that the human-commands mode is preferred over the human-requested mode, as the interactions with the robot are perceived as more efficient and natural, and the robot perceived as more intelligent. However, the higher standard deviation on the human-commands mode for intelligence can be explained by some participants finding the robot more intelligent for understanding commands, while others found the robot less intelligent for requiring commands to know what to do.

For the proactive and the autonomous mode the values are very similar. In our setting the participants often performed their action before the robot finished its action so that it was not necessary to perform an action proactively. As a consequence, three participants explicitly stated that they did not recognize any difference between the two modes while only one participant stated that they were impressed that the robot picked a block before it was possible to place it during the proactive mode. One participant suggested that the robot might use two arms at the same time to proactively pick a block while placing an other block. The task and instructions used in this study did not allow for discrimination between these modes in most runs. More complex tasks could create situations where the differences are more obvious.

We measured the task completion times for the individual modes. The proactive mode ($M = 55.71$, $SD = 9.36$) and the autonomous mode ($M = 59.00$, $SD = 17.19$) were the fastest, followed by the human-requested mode ($M = 79.14$, $SD = 21.41$) and the human-commands mode ($M = 81.00$, $SD = 45.09$). In reactive mode the task lasted the longest ($M = 116.00$, $SD = 12.54$).

General comments about the system included that the robot moved very slowly, only used one arm where it could have used two, and that it sometimes did not grasp the objects accurately. One participant queried what would happen if the participant did something wrong, such as placing an object incorrectly or even hiding an object. Another participant mentioned that they would like to have been able to ask the robot to finish off the task when there were still two actions remaining rather than needing to ask the robot to do something twice.

CONCLUSION AND FUTURE WORK

In this paper we investigated in human-robot collaboration during on-table tasks. We presented a system for on-table task execution and conducted a pilot study to compare different interaction modes. As in previous work by Baraglia et al [1] the participants preferred the proactive and autonomous mode over the reactive mode. We found out that for our basic task the proactive and autonomous mode is very similar. The small number of participants limit the findings of our study, as does the within subject design of the study, with possible effects from repeating the task with the robot 5 times. However, even with a small number of participants, we were able to identify several improvements for future investigation and development of our system.

For future work we plan to improve our system based on insights from the pilot study. Improvements for the general system could include better perception of the objects, increasing the speed of the grasp and place routines, using both arms for picking and placing objects concurrently, and providing feedback from the robot, for example, when an action requested by the human cannot be performed.

We plan to conduct a broader study with larger groups and more complex tasks to gain a deeper understanding in action-selection policies. The set of tasks will be designed to vary the amount of collaboration required to complete the task as the nature of the task may affect which action-selection policy is preferred.

ACKNOWLEDGMENTS

We thank the volunteers who participated in the study. We thank all current and previous members of the lab for the development of the code base that made the study presented in this paper possible.

REFERENCES

1. Jimmy Baraglia, Maya Cakmak, Yukie Nagai, Rajesh Rao, and Minoru Asada. 2016. Initiative in robot assistance during collaborative task execution. In *11th*

ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, 67–74.

2. Sandra Devin and Rachid Alami. 2016. An implemented theory of mind to improve human-robot shared plans execution. In *11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 319–326.
3. Bradley Hayes and Brian Scassellati. 2013. Challenges in shared-environment human-robot collaboration. In *Proceedings of the Workshop on Collaborative Manipulation, 8th ACM/IEEE International Conference on Human-Robot Interaction*, Vol. 8.
4. Séverin Lemaignan, Mathieu Warnier, E Akin Sisbot, Aurélie Clodic, and Rachid Alami. 2016. Artificial cognition for social human–robot interaction: An implementation. *Artificial Intelligence* (2016).
5. Bilge Mutlu, Allison Terrell, and Chien-Ming Huang. 2013. Coordination mechanisms in human-robot collaboration. In *Proceedings of the Workshop on Collaborative Manipulation, 8th ACM/IEEE International Conference on Human-Robot Interaction*. 1–6.
6. Stefanos Nikolaidis, Przemyslaw Lasota, Ramya Ramakrishnan, and Julie Shah. 2015. Improved human–robot team performance through cross-training, an approach inspired by human team training practices. *The International Journal of Robotics Research* 34, 14 (2015), 1711–1730.
7. Stefanos Nikolaidis, Swaprava Nath, Ariel D Procaccia, and Siddhartha Srinivasa. 2017. Game-Theoretic Modeling of Human Adaptation in Human-Robot Collaboration. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 323–331.
8. Ronald PA Petrick and Mary Ellen Foster. 2016. Using General-Purpose Planning for Action Selection in Human-Robot Interaction. In *2016 AAAI Fall Symposium Series*.
9. Andreas J Schmid, Oliver Weede, and Heinz Worn. 2007. Proactive robot task selection given a human intention estimate. In *The 16th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 726–731.
10. Julie Shah, James Wiken, Brian Williams, and Cynthia Breazeal. 2011. Improved human-robot team performance using chaski, a human-inspired plan execution system. In *Proceedings of the 6th International Conference on Human-Robot Interaction*. ACM, 29–36.