Studying Coordinating Behavior in Human-Robot Task Collaborations Using the PR2

Aaron St. Clair and Maja J. Matarić, *Fellow, IEEE*

***Abstract*—In order for a robot to coordinate its actions with people in collaborative task contexts the robot must be able to provide timely, informative feedback. This process requires the robot to infer the intent of its partners, assess the alignment of its planned actions with those of its collaborators, and if necessary issue communication actions that increase alignment. In order to test these capabilities in a repeatable, controlled environment, we have developed an augmented reality task simulation system that allows a person and a physical robot to carry out simulated tasks while being monitored by environmental sensors. This system is evaluated with a PR2 on a collaborative herding task.**

I. Introduction

As robots are capable of executing sequences of high-level tasks in household- or office-like environments it is necessary to consider how a robot should coordinate its actions with others. In human collaboration, people exhibit some reasonably well-understood behaviors that facilitate joint situational awareness[1]. It has also been demonstrated that a robot that coordinates its task execution by accounting for the requirements of a human collaborator is preferred to one that does not [2]. We hypothesize that a robot that communicates using natural social modalities, primarily embodied gestures, will further increase the performance of the human-robot team and the person's satisfaction with the interaction. In order to study team communication behavior, we have developed a task simulation system that enables the Willow Garage PR2 to collaborate with a user on a synthetic task in a collocated space.

II. Human-Human and Human-Robot Collaboration

Studies have demonstrated, during the course of a collocated, collaborative task, people adapt their speech and stage their actions in response to the person or people they are interacting with in order to establish alignment [3]. Alignment between collaborators reduces ambiguity by establishing consistent representations, concerning, for example, work objects or frames of reference [4]. This is manifested in task-level communication as lexical and syntactic alignment and a consistent vocabulary for issuing coordinating speech. In addition to speech, embodied gesture has also been shown the potential to be more effective than speech alone or the combination of speech and gesture for issuing instruction [5].

Prior work in human-robot interaction (HRI) has focused on coordinating actions during collocated collaboration through various means including intent recognition [6], [7]; learned spatiotemporal models of task performance [8], [9]; and through verbal communication such as turn-taking or action announcement [2], [10]. These systems in general attempt to incorporate Theory of Mind, in which mental state is attributed to other agents in the environment, with robot planning and control in order to avoid conflicting actions. Our work is concerned with integrating embodied social communication with task control by learning a flexible communication policy allowing a robot to issue communicative feedback such as deictic gesture [11] without assuming a specific communication structure *a priori*.

In order to be useful, the system should be capable of generalizing across different tasks that the robot may already know how to do on its own as well as multiple users or populations with different communication tendencies. The specific context of collaboration allows for the simplifying assumption that participants are working together to accomplish a shared set of goals. This enables the robot to make use of its own task controller or planner to evaluate and contextualize the actions of others via perspective-taking. A communication policy will be learned over time in a Theory of Mind-inspired state space constructed of each agent's current state as well as the estimated states of others as perceived by the robot. This approach will be fully detailed in a later paper.

III. Platform

A. Task Simulator

The approach is validated on a challenging cooperative task involving multiple people and robots who must communicate effectively to achieve a collaborative goal in a dynamic environment. In order to conduct repeatable experiments in several diverse task environments with human participants, we have developed an augmented reality task simulation system. The task simulator models the behavior of multiple virtual agents, static or dynamic, over time. The agents are projected onto the floor of the room via an overhead projection system. Simultaneously, the system makes use of environmental sensing to track people and robots in the same space, allowing them to interact with the virtual agents through physical action. The environment is calibrated by assuming a rigid transformation from the virtual space to the physical space of the room, allowing for generation of augmented sensor data, such as laser scans,

Manuscript received August 5, 2011. This work was supported in part by National Science Foundation (NSF) grants CNS- 0709296, IIS-0803565, and IIS-0713697 and ONR MURI program (N00014-09-1-1031).

A. St. Clair (corresponding author) and M. J. Matarić are with the Computer Science Department at the University of Southern California, Los Angeles, CA 90089-0781 USA (e-mail: astclair@usc.edu; mataric@usc.edu).

from the point of view of a physical robot in the space. The PR2 was selected as a test platform due to its robust sensing, allowing for accurate localization and people tracking, and its omnidirectional maneuverability, which is useful for avoiding obstacles without having to rotate in place.

This environment has several advantages for conducting tasks compared with using physical objects and/or confederate experimenters. First, it allows for many repeatable, dynamic agents that move autonomously in a directed or pseudo-random manner. It also allows for tuning the velocity, shape, and behavior dynamics of the simulated agents to make a given task easier or more difficult as necessary. Finally, since it does not depend on physical objects, several different tasks can be conducted quickly by switching the task controller. The simulator abstracts away physical tasks that we do not focus on, such as object manipulation, allowing us to focus on collaborative behavior. This is consistent with the notion of "research tasks" described by Martin et al. [12] as a systematic abstraction of a real-world task. Despite this abstraction, we are able to preserve some of the complexity of real-world environments, such as partial observability, noise, and occlusion by augmenting the robot's sensor data. We feel that this represents a reasonable tradeoff of realism for repeatable HRI experiments.

B. *Task-Relevant Social Communication on the PR2*

In order for the PR2 to collaborate with people effectively, it must be able to issue and receive social feedback while performing a task. The gestures implemented on the robot as well as those it is capable of recognizing will be informed from human-human pilot experiments in addition to literature on human behavior [13]. As we anticipate the communication dynamics to vary somewhat depending on the command hierarchy and personalities of the two participants. To control for this, we will be administering personality surveys and pairing participants so that they are either matched (undergraduates of similar age) or unmatched (undergraduate and graduate student).

To enable the PR2 to issue social feedback while performing a task, we have implemented a deictic gesture system allowing the PR2 to point to locations in space with its head or arms. The system incorporates collision avoidance for the arms in order to prevent hitting people or objects in the environment during gestures. In order to receive social feedback from others we will employ a camera-based laser tracking system allowing people to indicate points in the environment to the robot using a hand-held laser pointer. Finally, we plan on using the people-tracking provided by the Microsoft Kinect™ mounted on the PR2's head to sense natural social gestures when visibility allows.

IV. Experimental Design

We will validate the performance of the system in a series of human participant experiments. The experiments will initially feature a single robot and one person performing a task. The full communicative system will be compared to the robot alone, the person alone, and the robot and person working in parallel i.e., the robot will execute a state-based planner and will not issue or respond to feedback from the person. The first task controller that we have implemented simulates a herding task in which a group of virtual agents leave a central herd at random and must be collected and returned by the physical agents. Since realistic herd simulations are not our main interest, the herding dynamics are greatly simplified by allowing a single physical agent to capture and herd any single loose virtual agent. The primary focus of the collaboration in this scenario will be allocating herding responsibility and notifying partners of other un-herded agents. The accompanying video demonstrates the task simulation with the PR2 performing the task and herding the virtual agents. A series of experiments involving collaborative human-robot interactions and different tasks are planned and in progress.

References

[1] E. Entin and D. Serfaty, "Adaptive team coordination," Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 41, no. 2, p. 312, 1999.

[2] J. Shah, J. Wiken, B. Williams, and C. Breazeal, "Improved human-robot team performance using chaski, a human-inspired plan execution system," in Proceedings of the 6th International Conference on Human-Robot Interaction, pp. 29–36, ACM, 2011.

[3] S. Whittaker, "Things to talk about when talking about things," Human– Computer Interaction, vol. 18, no. 1, pp. 149–170, 2003.

[4] M. Pickering and S. Garrod, "Toward a mechanistic psychology of dialogue," Behavioral and Brain Sci., vol. 27, no. 02, pp. 169–190, 2004.

[5] S. Lozano and B. Tversky, "Communicative gestures facilitate problem solving for both communicators and recipients," Journal of Memory and Language, vol. 55, no. 1, pp. 47–63, 2006.

[6] A. Thomaz and C. Breazeal, "Teachable robots: Understanding human teaching behavior to build more effective robot learners," Artificial Intelligence, vol. 172, no. 6-7, pp. 716–737, 2008.

[7] C. Breazeal, J. Gray, and M. Berlin, "An embodied cognition approach to mindreading skills for socially intelligent robots," The International Journal of Robotics Research, vol. 28, p. 656, May 2009.

[8] C. Crick and B. Scassellati, "Inferring narrative and intention from playground games," Proceedings of the 7th IEEE International Conference on Development and Learning (ICDL'08), pp. 13–18, August 2008.

[9] G. Hoffman and C. Breazeal, "Effects of anticipatory perceptual simulation on practiced human-robot tasks," Autonomous Robots, vol. 28, pp. 403–423, May 2010.

[10] 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, vol. 35, no. 4, pp. 460–470, 2005.

[11] A. St. Clair and M. J. Matarić. "Task Coordination and Assistive Opportunity Detection via Social Interaction in Collaborative Human-Robot Tasks ". In 2011 International Conference on Collaboration Technologies and Systems Workshop on Collaborative Robots and Human Robot Interaction (CTS-CR-HRI), Philadelphia, PA, USA, May 2011.

[12] E. Martin, D. R. Lyon, and B. T. Schreiber, "Designing synthetic tasks for human factors research: An application to uninhabited air vehicles," In Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, pp. 123-127. Santa Monica, CA, 1998.

[13] J. Cassell, Y. Nakano, T. Bickmore, C. Sidner, and C. Rich, "Non-verbal cues for discourse structure," in Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, pp. 114–123, Association for Computational Linguistics, 2001.

[14] R. Kraut, S. Fussell, and J. Siegel, "Visual information as a conversational resource in collaborative physical tasks," Human-computer interaction, vol. 18, no. 1, pp. 13–49, 2003.