Gesture, Gaze, Touch, and Hesitation: Timing Cues for Collaborative Work

Justin W. Hart
Department of Mechanical Engineering
University of British Columbia
justin.hart@mech.ubc.ca

Brian Gleeson
Department of Computer Science
University of British Columbia
brian.gleeson@gmail.com

Matthew Pan
Department of Mechanical Engineering
University of British Columbia
matt.pan@gmail.com

AJung Moon
Department of Mechanical Engineering
University of British Columbia
ajungmoon@gmail.com

Karon MacLean
Department of Computer Science
University of British Columbia
maclean@cs.ubc.ca

Elizabeth Croft
Department of Mechanical Engineering
University of British Columbia
elizabeth.croft@ubc.ca

ABSTRACT

When multiple agents interact in order to perform a collaborative task, conflicts will arise over access to shared resources or when one agent's ability to act relies on the actions of another agent. In such cases, even perfectly cooperative agents will encounter problems due to imperfect knowledge of each other's behavior, requiring them to communicate with each other. The CHARM project seeks to develop robot assistants which work alongside human workers in a manufacturing environment. Towards this goal, we have studied nonverbal cues for timing coordination between human collaborators. We have modeled these cues for use with robots and software systems and validated them through human-robot interaction studies. This paper provides an overview of these studies, as well as a high-level description of a system currently under development which is intended to allow a robot to model the time required for individual human workers to complete the tasks to which they are assigned and to adapt its timing and work pace to match that of its human collaborators.

1. INTRODUCTION

When multiple agents collaborate to perform a task, temporal and ordering conflicts arise. Coordination is required in various scenarios: where agents must share a resource, when physically embodied actors want to occupy the same space, or when one agent's tasking relies on the completion of a task by another agent. The determination of an optimal plan that coordinates the efforts of multiple agents is part of an active area of computer science and operations research known as planning and scheduling. The output of planning software is a plan describing factors such as the order in which tasks should be performed.

If a planner is able to plan the actions of every actor collaborating on a task, then it is able to resolve potential conflicts by instructing their behavior, communicating to an actor to cede right-of-way to another, or dictating that a task be completed by a certain time. Difficulties arise when the planner is required to work with imperfect knowledge regarding the behavior of agents that are not directly under its control. This scenario has an analog in human behavior, since even people who wish to fully cooperate on a task do not have perfect insight into each other's mental states. Humans utilize a number of cues that communicate their intentions to each other through implicit cues such as body language and visual attention, and explicit cues such as vocalizations and gesturing. In the Collaborative Human-Focused Assistive Robotics for Manufacturing (CHARM[1]) project, we have been studying the suitability of non-verbal cues for use by robot assistants in manufacturing scenarios.

In this paper, we first overview a series of studies we have conducted to explore the effectiveness and design of human-inspired, non-verbal communicative cues (both robot-human and human-robot in a variety of human-robot interaction (HRI) structures). These studies demonstrate the potential of non-verbal cues to communicate and coordinate task timing during paired collaborative tasks. We then present our plans for a system that utilizes these communicative cues to infer and adapt to the timing and task performance of a human operator in a human-robot collaborative assembly task context.

2. COMMUNICATIVE CUE STUDIES

Fluent timing and teamwork in human-robot collaborative tasks requires an understanding of the turn-taking and task-flow regulation cues used by humans and how these cues can be applied to HRI. In a series of studies on turn-taking, timing, and task flow we have observed human-human pairs in collaborative tasks and have adapted our findings from these observations to HRI. We focused on non-verbal cues

[1]CHARM is a collaboration between investigators at University of British Columbia, Université Laval, and McGill University.

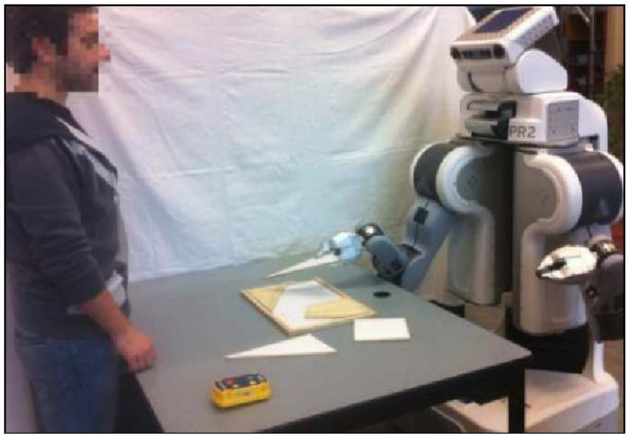

Figure 1: Communicating the timing of turn exchange using postural changes in a human-robot task.

in consideration for industrial contexts where verbal communication may be impractical or undesirable.

Our studies encompass a range of interaction structures. In Sections 2.1 and 2.2, we summarize work on turn-taking and task-flow control. These studies focus on tasks with structured turns where task control alternates between partners. Section 2.3 describes studies with structured turns, but with the human maintaining primary control of the task. Section 2.4 summarizes a co-manipulation task (handover) with simultaneous motion and shared task control. In Section 2.5, we discuss an asynchronous collaborative task, where partners share a workspace but have no structured turns. By investigating a wide range of interaction structures, we aim to contribute to the foundational knowledge for general human-robot turn-taking and temporal coordination.

2.1 Non-Verbal Cues for Turn-Taking

In collaborative tasks that include turn-taking, it is essential for the involved agents to be able to communicate when a turn is given or taken by another agent in the interaction. In a study involving 18 participants, we observed and identified postural gestures during turn-taking interactions [1]. Participants, formed into 15 unique dyads, stood across a table and collaborated to complete tangram puzzles without verbal communication. We required that participants alternate placing pieces into the puzzle solution and observed the non-verbal cues they used to regulate turn-taking.

Our findings show that while some participants occasionally used explicit hand gestures, participants generally signaled the end of their turn by changing posture. Common postural signals included participants placing their hands down at their sides, placing their hands on the table, or stepping back from the table. Often, we found that participants used two of these cues in conjunction with each other. In this study, with no structured timing of turns, results showed that participants were able to effectively indicate the timing of turn exchanges through changes in body posture. In ongoing work, we are studying if and how similar cues can be used in human-robot interactions involving similar unstructured turn-taking tasks (Figure 1).

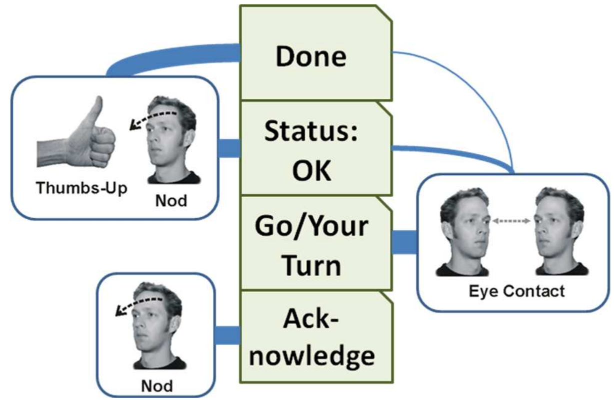

Figure 2: Common flow regulation gestures in a collaborative assembly task. Most gestures had multiple meanings, depending on the task context. The width of the line linking gesture and meaning indicate the prevalence of that gesture-meaning pair in our study.

2.2 Cues for Task Flow Regulation

We have found that people use different types of non-verbal cues to regulate the flow of a task when a collaborative task does not have a fixed pattern of turn-taking. In a study of 16 participants, pairs of participants worked together to complete simple simulated assembly tasks, assembling stacks of flat wooden geometric figures and placing them into pre-assigned locations on a board. Participants were given different task roles and knowledge (e.g., only one participant was allowed to fetch parts from the supply, only one participant knew the assembly order, etc.). Timing of the turn-taking interaction was unstructured and unplanned. In an effort to focus communication through hand gestures, the participants donned sunglasses and a face mask in order to hide their eyes and mouths from view of their collaborating partner.

We observed two modes of communication during the experiment: part manipulation communication - which involved conveying the placement and movement of parts; and flow regulation communication - where gestures were aimed at directing turn-taking, exchanges, and task flow. Our findings on part manipulation communication, along with a human-robot study, can be found in [6]; a brief summary of flow regulation is reported in [5].

When regulating task flow, participants exhibited a strong preference for gestures involving the head and face, despite our efforts to prevent facial communication. For example, participants nodded their heads to communicate that they had completed the current task, or turned to look at their partner in an attempt to establish eye contact to communicate "your turn". Figure 2 shows the most common gestures used during the study and their associated meanings.

While gestures involving gaze and head orientation are more difficult for a machine vision system to recognize, and impossible for non-anthropomorphic robots to execute, they allow human operators to regulate task flow quickly and intuitively, in most cases without taking their hands from their

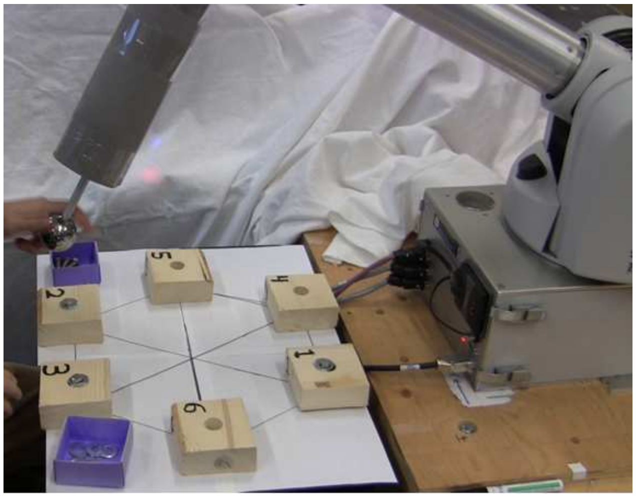

Figure 3: The user taps and pushes the robot to regulate turn-taking and task flow in a simulated assembly task.

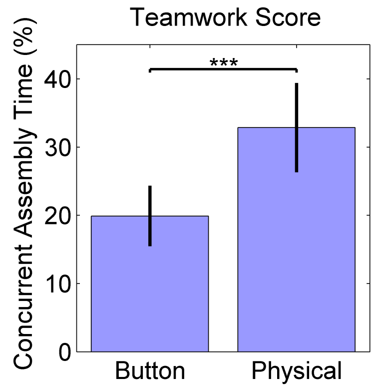

Figure 4: Direct Physical Control significantly improves human-robot teamwork fluency in a turn-taking task.

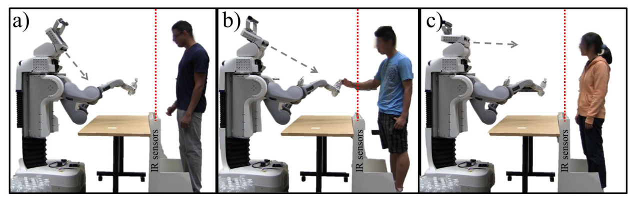

Figure 5: Gaze in human-robot handovers. During the handover, the robot either (a) remains looking down, (b) looks at the intended handover location, or (c) looks the receiver in the face. The position of the light curtain is shown by the red line. Participants completed handovers faster when the robot shifted its gaze to the handover location (b).

work. We are currently considering how these communication concepts could be adapted to a (headless) robot arm.

2.3 Turn-Taking and Task Flow Regulation using Touch

Direct physical commands issued by a user physically tapping or pushing a robot are a fast and effective way of providing explicit timing cues to a robotic system. While our current implementation of touch interaction is only one-way (human-to-robot), it has the advantage of allowing users to keep their attention, gaze and hands in the workspace at all times. In time-sensitive tasks, humans can execute direct physical commands quickly and robots can interpret the commands without significant processing time. Our work on direct physical control is currently in press [4].

We conducted a series of studies with 43 participants engaging in simulated collaborative assembly tasks with either a 7-DOF Barrett Whole Arm Manipulator (WAM) or a small desktop Phantom Omni robot. Participants regulated turn-taking and task advancement by tapping or pushing the robot as shown in Figure 3.

Compared to traditional push-button controls used to interact with robots, we found that direct physical interaction improved users' ability to control the flow of the human-robot collaboration in most of the experimental tasks, resulting in reduced task completion times, improved teamwork fluency (Figure 4), and a better overall user experience. While human-robot touch interactions are not appropriate for all tasks, we have found it to be a fast and intuitive communication channel that can improve task flow and overall performance in complex human-robot tasks.

2.4 Timing Cues from Gaze in Human-Robot Handovers

Object handovers are an important and necessary building block for physical human-robot collaborative interactions. Much like many other types of interaction between humans and robots, the quality of a handover is highly sensitive to timing. A successful, fluent handover requires that both giver and receiver agree on when the receiver should grasp the object, when each party has control of the object, and when the handover is complete. Failure to communicate and synchronize timings between giver and receiver is a frequent cause of handover failure [10]. In a study being presented at HRI 2014, we show that robot gaze using the head on an anthropomorphic robot can be used to influence human handover timing [9].

In our study, 96 naïve participants engaged in handovers with a Willow Garage PR2 robot. We designed three gaze patterns based on observations of human-human handovers. In each trial, the robot started by looking at the object in its hand, then shifted its gaze to one of the following locations: (a) no gaze, looking at the ground; (b) location gaze (or shared attention gaze), looking at the intended handover location; and (c) facial gaze, looking at the receiver's face (Figure 5). For each condition, we designed the gaze movements (head movements) to signal handover location and timing. Infrared sensors arrayed as a light curtain detected when a participant initiated a handover, while sensors in the robot hand and wrist detected when the human grasped the object and when the handover was complete.

We found that the robot's gaze influenced the timing of the human participant's reaching behavior. When the robot directed its gaze at the handover location, participants initiated their reach sooner and completed the handover faster. We hypothesize that the direction of the gaze communicated the intended handover location, and that the timing of the gaze and hand motions communicated the intended handover timing. Furthermore, we suspect that the gaze provided a social cue focusing the participant's attention on to the handover location. Our results show that robot gaze can be used to improve timing and fluency in handovers, and suggest that gaze may be used to control timing in other physical human-robot interaction.

2.5 Resolving Resource Conflicts using Hesitation

Resource conflicts are common in asynchronous collaborative work, arising when two parties simultaneously reach to the same location. Human-human pairs resolve these conflicts non-verbally using hesitation gestures, halting hand motions that communicate an awareness of the conflict and can be used to yield right-of-way. In fast-paced human-robot tasks, hesitation gestures could be used as a subtle, intuitive means of negotiating access to shared resources and resolving problems in reach timing. Our work on human and robot hesitation gestures is reported in [7, 8].

In a human-human study, a pair of male undergraduate students were instrumented with a pair of two inertial sensors attached to each of their right arms. Seated across from each other at a small table, they were instructed to perform a simple task which required access to a shared resource, touching a sponge placed in the middle of the table when prompted by audible tones through a pair of headphones that each was wearing. Based on recordings of their arm motion during this study, we designed a robot hesitation motion trajectory, Figure 6.

For an initial experiment, we recorded videos of a researcher and a 6-DOF CRS A460 robot engaging in a collaborative reach-and-retract task with occasional resource conflicts. The robot responded to conflicts (potential collisions) with (a) blind motion, i.e., collision, (b) a 'robotic' trapezoidal collision avoidance trajectory, or (c) our designed human-like hesitation gesture for collision avoidance. In a video survey of 58 participants, we found that the designed robot hesitation gesture successfully communicated a state of hesitation and yielding and improved the perceived anthropomorphism of the robot compared to the 'robotic' trajectory.

In a second study testing the same three responses to resource conflicts, 24 participants engaged in a fast-paced reach-and-retract task with a 7-DOF Barrett WAM. Both human and robot asynchronously reached for a shared container of parts, resulting in occasional simultaneous reaches. As before, the designed robot hesitation trajectory successfully communicated hesitation and yielding, but the results were unclear on how hesitation affected the users' perception of the robot or task completion timing.

Our initial results show users understand these robotic hesitation gestures, but more work is required before hesitation

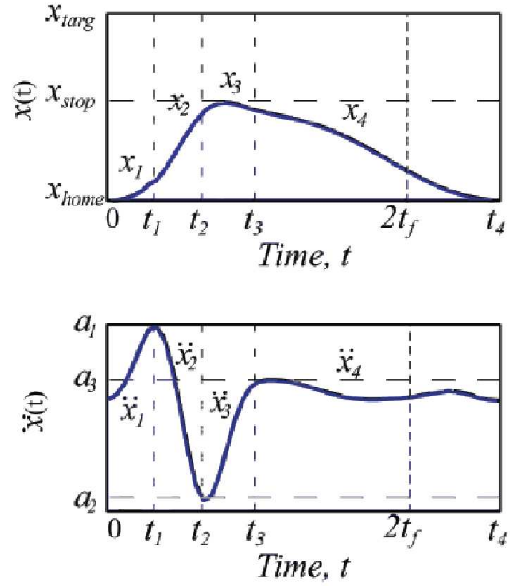

Figure 6: Human-like hesitation gesture trajectory in position (top) and acceleration (bottom). Ratios between time and acceleration parameters ($t_i : a_i$) define the gesture. These ratios are taken from observations of human hesitations. This trajectory successfully communicates hesitation and yielding behavior to human partners.

can be used to regulate timing and resources access in collaborative tasks. Between humans, hesitation is a two-way communication channel, with partners engaging in a short gestural negotiation to resolve conflicts. In ongoing work, we are investigating bidirectional human-robot hesitation negotiations as well as studying other variations on the hesitation gesture.

3. NEXT: ADAPTING TO HUMAN TIMING CUES

Traditional factory automation robots do not interact with humans during their operation and, rather, operate in carefully-controlled and designed environments. Alternatively, the CHARM project addresses the development of robot assistants that work alongside human workers. These assistive robots are intended to work collaboratively with humans, as opposed to performing tasks entirely autonomously (as in the case of a traditional factory automation) or fully under the control of the human operator (as in the case of a tool). We aim to develop human-robot interactions that are natural for workers who may not be experts in robotics. Such interactions should be fluid and timely, minimizing the time that the worker spends on instructing the robot, and maximizing the time spent performing assembly tasks. This means having the robot perform its tasks at the right time with respect to the human collaborator; not making the worker wait by being too late, and not getting in the way by being too early. Furthermore the robot should correctly respond to the worker's cues to speed up, slow down, or get out of the way, and adapt to both changing user preferences and differences between workers.

To this end, we intend to attach inferred timings to a model of a collaborative assembly task performed by a human worker and a robotic assistant. These timings will be based on real-world performance of the task by both the human worker and the robotic assistant, and inferred online during operation by monitoring task performance. The system will learn the pace at which the human worker performs their portions of a collaborative assembly task, computing timing statistics through repeated iterations of the task. In addition, the worker will be able to provide input to the robot regarding their preferences, (e.g. please speed up/slow down) using communicative cues, such as those outlined in Section 2. Through the use of such modeling and cues, we intend to design a system that adapts to the pace of the workers with which it interacts, allowing it to behave less like a tool and more like a collaborative member of a human-robot team.

3.1 Design Goals

By using existing planning software and modeling techniques, we hope to minimize the amount of effort placed on infrastructure, and instead focus on the HRI aspects of this project. Preliminarily, we plan to monitor the worker's progress on the collaborative task through implicit cues such as the tool they are using, the pose of the tool with respect to the assembly, and their position in the workspace. We plan to integrate our work on nonverbal communicative cues into the system, allowing the robot to perform actions such as hesitation when unexpected resource conflicts arise. Explicit feedback will be provided to the system through both non-contact gestures such as waving the robot in to request that it speed up, as well as contact gestures such as pulling and pushing on the robot in order to instruct it.

Once timings have been incorporated into our model, we will be able to use the model to optimize a number of timing related criteria. For instance, the same model should be useful for optimizing not only the overall time it takes to complete the assembly task, but also to minimize idle time and maximize concurrent productive work. An ideal plan should minimize annoyances such as interruptions, and allow efficiently respond to error conditions. These metrics may come into conflict, so trade-offs and overall user satisfaction will be evaluated through the use of a post-interaction survey in order to guide our design.

3.2 High-Level Design

Enabling a robot to perform a collaborative assembly task with a human worker is partially facilitated by the existence of precise instructions for manufacturing assembly. These instructions can be rewritten in a high-level planning language (such as PDDL [3] or STRIPS [2]), which represent problems and tasks as a set of actions that can be taken by the robot. Each action has a list of pre-conditions that must be met in order before the action taken (e.g., there must be a bolt on the work piece before the robot can attempt to fasten it) and post-conditions that result from taking this action (e.g. fastening the bolt mates two separate parts). Internally, planners chain these actions together to link a starting state to a goal state. In our system, where the planner will be run in iterative cycles, the starting state for each cycle will be derived from sensor data, whereas the goal state will be the completed assembly task. Many planning languages are able to account for factors such as the expected time for completion of an action, allowing planners to optimize resource allocation and the amount of time to accomplish a task. These abilities will allow us to attach estimates of task times based on real-world performance to each action, in turn allowing us to optimize the timing of events in the human-robot interaction. Once implemented on the robot, the same planning language can be used to model both the expected behavior of the human collaborator and the actions to be taken by the robot.

Knowledge regarding the environment, the human operator, and the current state of the task will be obtained through the use of state-of-the-art modeling and sensing software developed by collaborators at our partner institutions on the CHARM project. In the current implementation of our environment awareness system, point cloud data is sampled through a set of four Microsoft Kinect 3D sensors and processed to produce a world model representation. The robot uses this representation to identify and reason about operators and objects in its environment. A skeleton tracker, developed by Andrew Phan at McGill University, will be used to track the body pose of the human collaborator, allowing us to interpret gestures and make inferences about the task state based on the worker's motion and posture. Information from this skeleton tracker will be complimented by object tracking and workspace occupancy data, provided by systems developed by Denis Ouellet and Dominique Beaulieu, respectively at Université Laval. Both of these systems will provide inputs to the Situational Awareness Database (SADB), developed by Olivier St-Martin Cormier at McGill University. The SADB is used to provide an overview of the current world state, as estimated from data provided by the various inputs to the system including sensors, a model of task instructions, and prior performance fed back from previous iterations of the task activity.

By integrating data from the SADB, our system will be able to reason about the current world state as it relates to its assigned assembly task. This will allow the robot to track the human's progress on the task, reason about the human's actions, and act appropriately based on the planner's output.

3.3 Evaluation Plan

To evaluate the system, we plan to exercise a two-phase evaluation.

3.3.1 *Experimental Evaluation*

The first phase is a user study at the CARIS Laboratory at UBC. Participants will be recruited from the body of college students, and the task will be a highly repetitive simulated assembly designed to quickly and thoroughly evaluate the system.

Participants will take part in a bolt insertion task similar to that discussed in Section 2.3 of this paper. They will be instructed to insert bolts into wooden blocks in a predesignated pattern, with a 7-DOF Barrett WAM following from behind to 'secure' the bolts that they have inserted into the wooden blocks (gestured by lowering the robot's end-effector over the top of the bolt). The timings will be designed to insure that slow behavior on the part of the robot will protract the amount of time required to perform the task, whereas fast timing (e.g., moving into position to perform the next part of the task too early) will cause the robot to either collide with the human participant (at a slow, safe speed) or to obstruct the participants task area or path of their arm motion. Participants will repetitively perform this bolt-assembly task until the system's model of the speed with which they perform the task converges with the participant's.

Users will be presented with a post-interaction survey in order to help to evaluate their experience using the system.

3.3.2 Practical Validation

Annually, the collaborators on the CHARM project (UBC, Université Laval, McGill University) meet at Laval for an integrated demonstration of the technologies developed over the previous year. Among the researchers involved in the CHARM project, three Subject Matter Experts (SMEs) will be invited to provide input regarding the real-world viability of the system.

In this study, the SMEs will perform a car door assembly task, inserting components into a car door in an assembly area wherein a robot will assist them with this task. A custom robot consisting of a robotic arm mounted on a mobile gantry platform assists in this task simply by presenting the next part to be inserted into the door to the operator, after having selected it from a variety of possible components on a shelf. As this experimental task takes longer to perform than the bolt insertion task and can only be performed on-site at the Université Laval (due to unique hardware requirements demanded by the experiment), we will rely on operator interviews and commentary rather than computed statistics to gain additional insights into the performance of the system.

4. EXPECTED CONTRIBUTIONS

The CHARM project has made significant contributions to HRI through the use of implicit and explicit nonverbal cues communication between humans and robot assistants. This has been accomplished in part by modeling these cues as used by humans in human-human interactions, and adapting them for use with robots.

In this work, we will perform a similar modeling process online, by modeling and adapting to the performance of workers collaborating with the robot in a collaborative assembly task. Though this work is still in the planning stages, we expect for the system to be able to adapt to the preferred timing and work pace of individual workers, providing an experience that is suitable for both novices, who may work more slowly and require more help, and advanced users, who may work very quickly, not wanting to wait for the robot to catch up. Constructing the system in this way will allow us to perform detailed analyses of task timing and performance by collecting statistics regarding individual portions of the task and building models of worker performance. Rather than attempting to learn an optimal behavior, we will attempt to model task performance, using this as a means to plan optimal behavior. We hope that this will allow us to performed detailed HRI studies regarding how a robot can adapt its timing and behavior to the work patterns of a human collaborator.

5. ACKNOWLEDGEMENTS

The authors would like to recognize and thank all of our collaborators on the CHARM project. This work is supported by funding from General Motors and NSERC.

6. REFERENCES

[1] E. Calisgan, A. Haddadi, H. F. M. V. D. Loos, J. A. Alcazar, and E. A. Croft. Identifying nonverbal cues for automated human-robot turn-taking. In *Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication*, pages 418–423, Paris, France, September 2012. IEEE Computer Society.

[2] R. Fikes and N. Nilsson. Strips: a new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2(3–4):189–208, 1971.

[3] A. Gerevini and D. Long. Preferences and soft constraints in pddl3. In *Proceedings of the ICAPS-2006 Workshop on Preferences and Soft Constraints in Planning*, pages 46–53, Cumbria, UK, June 2006.

[4] B. Gleeson, K. Currie, K. MacLean, and E. Croft. Tap and push: Assessing the value of direct physical control in human-robot collaborative tasks. *Journal of Human-Robot Interaction*, in press.

[5] B. Gleeson, K. MacLean, , E. Croft, and J. Alcazar. Human-robot communication for collaborative assembly. In *Proceedings of GRAND 2013, the 4th Annual Conference on Graphics, Animation, and New Media*, Toronto, Canada, May 2013.

[6] B. Gleeson, K. MacLean, A. Haddadi, E. Croft, and J. Alcazar. Gestures for industry: intuitive human-robot communication from human observation. In *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction*, pages 349–356, Tokyo, Japan, March 2013.

[7] A. Moon, C. A. C. Parker, E. A. Croft, and H. F. M. Van der Loos. Did you see it hesitate? - empirically grounded design of hesitation trajectories for collaborative robots. In *Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1994–1999, San Francisco, CA, USA, September 2011.

[8] A. Moon, C. A. C. Parker, E. A. Croft, and H. F. M. Van der Loos. Design and impact of hesitation gestures during human-robot resource conflicts. *Journal of Human-Robot Interaction*, 2(3):18–40, January 2013.

[9] A. Moon, D. M. Troniak, B. T. Gleeson, M. K. X. J. Pan, M. Zheng, B. A. Blumer, K. E. MacLean, and E. A. Croft. Meet me where iâĂŹm gazing: How shared attention gaze affects human-robot handover timing. In *Proceedings of the 9th ACM/IEEE International Conference on Robot-Human Interaction*, Bielefeld, Germany, March 2014.

[10] K. W. Strabala, M. K. Lee, A. D. Dragan, J. L. Forlizzi, S. Srinivasa, M. Cakmak, and V. Micelli. Towards seamless human-robot handovers. *Journal of Human-Robot Interaction*, 2(1):112–132, 2013.