CADENCE for Collaboration and Companionship with Robots

Crystal Chao and **Justin Smith** and **Andrea L. Thomaz**
School of Interactive Computing
Georgia Institute of Technology
801 Atlantic Drive, Atlanta, Georgia 30332
{cchao, jsmith41}@gatech.edu, athomaz@cc.gatech.edu

Abstract

We give an overview of how the CADENCE architecture addresses the problem of turn-taking in embodied interaction.

Introduction

The vision of social robotics encompasses such everyday roles for robots as butlers, factory teammates, schoolteacher assistants, and information agents in public spaces. Each role lends unique requirements for social dominance and distance. A schoolroom or babysitting robot may exert authority over children but defer to parents and teachers. A home healthcare robot requires a different level of user familiarity than a robot receptionist.

To achieve such tailored interaction styles with a tractable amount of effort, it is important to develop general-purpose social cognition and behavior that work for a range of social situations. One such core social skill is *turn-taking*. Turn-taking often refers to the exchange of the conversational floor, but more broadly, it describes the exchange of any resources within a joint activity. Implicit within any social role is a turn-taking style that is appropriate for and effective in the performance of that role.

We are developing an architecture called CADENCE to control such turn-taking styles in social human-robot interactions. We aim for CADENCE to be applicable to domains involving multimodal human-robot *collaboration*, including collaborative dialogues. Collaboration requires careful initiative-balancing and transparent social cues in order for participants to reach common ground and fulfill their goals. In addition, we seek to apply CADENCE to improving *companionship*, specifically by producing *playful* turn-taking behavior. We contend that a robot with the creative capacity for computational play can establish greater rapport and social connection in a companionship role.

In this abstract, we give a brief overview of how CADENCE supports turn-taking in multimodal collaborative dialogues. Then we describe its application to two interaction domains that are under active development. One is a task that involves collaborative problem-solving and manipulation. The other is an object play setting in which the robot dynamically switches between leader and follower roles.

Controlling Reciprocal Social Interactions

To build robots that work and play with humans, we must address the embodied real-time nature of reciprocal social interactions. Turn-taking is a process of dynamically managing shared resources in an interaction. In an embodied dialogue with a robot, this can include the conversational floor, objects, and space. Resources are required to execute temporally extended actions, so agents who own resources at any given time have elevated control over the interaction outcome. If a robot cannot use embodied resources appropriately, goals will be achieved less successfully and social benefits will be lost.

The notion of turn-taking is of growing interest to those conducting research in HRI, spoken dialogue, and virtual agents. Spoken dialogue systems have previously formulated turn-taking in terms of minimizing system barge-ins while optimizing task success and completion times. For embodied agents, increasing importance is also being placed on the roles of nonverbal cues, such as eye gaze and gesture (Cassell and Thorisson 1999; Mutlu et al. 2009). Currently, it is common for interaction control to be modeled using a state-based representation such as a finite state machine (FSM) (Raux and Eskenazi 2009) or a partially observable Markov decision process (POMDP) (Williams and Young 2007; Rosenthal and Veloso 2011) and solved as a purely sequential problem. In these formulations, a state tends to correspond to a single turn, and only one state is executed at any time. This gives rise to the command-like, stop-and-go turn-taking structure we have come to expect from computational systems. More recent advances take the view of interaction as a problem of incremental processing, which produces more quick and naturalistic turn-taking timing (Schlangen and Skantze 2011).

Our research focus is to develop timing models and autonomous control that enable better turn-taking interactions. We take a cyclic approach to the problem in which we identify specific turn-taking phenomena in data collected from dyadic face-to-face interactions of humans with our robot; design and implement model-controllers based on our observations; and evaluate the success of the robot's turn-taking control in a subsequent user study. Thus far we have demonstrated: (1) that information flow and resource constraints are together the best predictor of floor passing in a reciprocal interaction (Chao et al. 2011); (2) that using an autonomous controller that actively yields the floor leads to a better balance of control, resulting in perceptions that the robot is a better teammate (Thomaz and Chao 2011; Chao and Thomaz 2012); and a turn-taking controller that can be parameterized to achieve different social dynamics (Chao and Thomaz 2013). Moving forward, we intend to apply our turn-taking framework to the problems of generating collaborative and playful behaviors with appropriate social dynamics and timing.

CADENCE

The Control Architecture for the Dynamics of Embodied Natural Coordination and Engagement (CADENCE) is a framework for autonomously controlling the multimodal behavior and turn-taking for a social robot. Because of the concurrent and real-time nature of interaction resource management, CADENCE is implemented using timed Petri nets (TPNs), which serve as a unified representation for modeling, control, and simulation. The design of CADENCE addresses the requirements of fluent turn-taking threefold:

1. *Interruptible modality actions* – To achieve fluent turn-taking, a robot requires the ability to interrupt its own actions. This could be because a communicative goal is achieved midway through the action execution or a contested resource is being yielded to the human. Thus, all action modalities in CADENCE are modeled as interruptible processes.
2. *Resource monitoring* – Embodied interactions are conducted over multiple resource types, such as the speaking floor, physical objects or space, and the robot's degrees of freedom. A resource monitor serves as the interface between processes competing to use these resources. The processes for the robot's actions request and release resources when executing, and perceptual transitions in the user model detect when resources become occupied or available through external events. Whenever resources are shared between the human and the user, the resource monitor acts as a turn-taking model with parameters that can be adjusted according to the social situation. CADENCE supports parameter settings for controlling relative resource usage, action spacing, interrupts, and tolerance for overlap or for resources going unused.
3. *Dynamic scheduling supervisor* – Intentions in CADENCE are represented as hierarchies of primitive actions, with temporal constraints between the actions. Actions also support disjunctive resource usage; for example, an object might be referred to using a speech-only referring expression or by saying "this" with a deictic gesture. To schedule resources correctly in arbitrary situations while allowing action concurrency for a queue of such intentions, CADENCE uses a dynamic scheduling backend called Tercio (Gombolay, Wilcox, and Shah 2013). Resources are allocated by balancing social reward with the schedule makespan. When resource availability changes, for example due to the human interrupting the robot, scheduling updates are performed in real time.

Our work with CADENCE focuses on the resources and actions required to support interactions using speech, eye gaze, gestures, and object manipulation. Examples of past turn-taking interactions that we have demonstrated on our humanoid robot Simon include playing the imitation game Simon Says (Chao et al. 2011), collaborating to solve the Towers of Hanoi (Chao and Thomaz 2012), and learning about tabletop objects (Chao and Thomaz 2013).

This past work focused on identifying factors and core system competencies essential to turn-taking. We identified the need for a semantic interface between a general turn-taking model and the domain-specific contents of an interaction, preferably in which the semantics are recognized and executed as incremental units (sub-turn level) rather than corresponding one-to-one with a turn. In addition, we identified and implemented control policies for the skills of *seizing* and *yielding* the floor. Next, we highlight two domains under active development and the key turn-taking problems we expect to address in these domains.

Collaborative assembly

Currently we are applying CADENCE to the collaborative assembly problem of a human and a robot constructing a block model together. In our domain, each agent has partial information about the solution, and they must reach common ground through situated dialogue to build the model successfully.

In this setting, turn-taking is conducted over the speaking floor, shared objects, and the shared model workspace. The resource constraints and the absence of complete information both factor into when an agent needs to wait, listen, speak, or take an opportunity to parallelize manipulation, gesture, and speech. An agent needs to be able to interrupt actions that dynamically lose relevance or are found to be incorrect. In addition, it is important that the agent take initiative at appropriate times to offer new information or to repair task and perceptual state.

Leading and following in collaborative play

Many dyadic interactions involve an asymmetric dominance relationship, in which one participant acts as the leader and one as the follower. The leader takes more initiative in directing the interaction outcome, and the follower's turns and actions support the leader's intentions. In previous work, we defined sets of parameters to represent opposite extremes of turn-taking behavior. *Active* parameters consisted of lower lapse tolerance, shorter spacing between actions, a higher ratio of floor time, and the ability to barge in while resisting interruption; *passive* parameters did the opposite.

However, in many social contexts, such roles may not be static. In collaborative play, the interaction partners are peers and trade roles throughout. Thus, a completely dominant or completely passive setting of the CADENCE parameters would not produce the best play companion. Instead, we hypothesize that the appropriate behavior is to switch dynamically between active and passive parameters. This requires the ability to recognize appropriate opportunities for role-switching as well as to monitor and control timing dynamics incrementally or in a context-sensitive way rather than only over the entire interaction duration.

References

Cassell, J., and Thorisson, K. R. 1999. The power of a nod and a glance: envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence* 13:519–538.

Chao, C., and Thomaz, A. 2012. Timing in multimodal reciprocal interactions: Control and analysis using timed Petri nets. *Journal of Human-Robot Interaction* 1(1):4–25.

Chao, C., and Thomaz, A. 2013. Controlling social dynamics with a parametrized model of floor regulation. *Journal of Human-Robot Interaction* 2(1):4–29.

Chao, C.; Lee, J. H.; Begum, M.; and Thomaz, A. 2011. Simon plays Simon says: The timing of turn-taking in an imitation game. In *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 235–240.

Gombolay, M. C.; Wilcox, R.; and Shah, J. 2013. Fast scheduling of multi-robot teams with temporospatial constraints. In *Proceedings of Robotics: Science and Systems (RSS)*.

Mutlu, B.; Shiwa, T.; Ishiguro, T. K. H.; and Hagita, N. 2009. Footing in human-robot conversations: How robots might shape participant roles using gaze cues. In *Proceedings of the 4th ACM/IEEE Conference on Human-Robot Interaction (HRI)*.

Raux, A., and Eskenazi, M. 2009. A finite-state turn-taking model for spoken dialog systems. In *Proceedings of Human Language Technologies (HLT)*.

Rosenthal, S., and Veloso, M. 2011. Modeling humans as observation providers using POMDPs. In *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 53–58.

Schlangen, D., and Skantze, G. 2011. A general, abstract model of incremental dialogue processing. *Dialogue and Discourse* 2(1):83–111.

Thomaz, A., and Chao, C. 2011. Turn-taking based on information flow for fluent human-robot interaction. *AI Magazine Special Issue on Dialogue With Robots* 32(4):53–63.

Williams, J., and Young, S. 2007. Partially observable Markov decision processes for spoken dialog systems. *Computer Speech and Language* 21(2):231–422.