Real-Time Changes to Social Dynamics in Human-Robot Turn-Taking

Justin S. Smith, Crystal Chao and Andrea L. Thomaz

Abstract— **In order for robots to work alongside humans in a range of domains, they will need to operate with a variety of social dynamics that each context will require. This paper builds on previous work with a parameterized turn-taking model, CADENCE, in which different parameter settings resulted in different social dynamics. In contrast to the static parameter settings of previous work, we now investigate the problem of changing these turn-taking parameter sets dynamically within a single interaction session. This ability is necessary for successful peer-to-peer collaborations, in which balance of control between leading and following must be maintained. We present our dynamic switching approach and an experiment with 15 participants. Our results confirm that it is possible to achieve the same changes in social dynamics within a single interaction session that were previously seen only between independent sessions of different parameter settings. Moreover, we show that such a change in social dynamics is contingent upon changing parameters at socially appropriate turn boundaries.**

I. Introduction

The vision of service robotics encompasses such everyday roles for robots as personal butlers, factory teammates, schoolteacher assistants, and information agents in public spaces. Each of these roles lends particular requirements to the nature of a successful interaction in terms of 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, it is important to develop general-purpose social cognition and behavior that work for a range of 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, the Control Architecture for the Dynamics of Embodied Natural Coordination and Engagement (CADENCE), to control such turn-taking styles in social human-robot interactions. 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 [3], 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. An experiment showed that these different parameterizations resulted in significantly different social dynamics between the human and the robot.

College of Computing, Georgia Institute of Technology, Atlanta, GA 30332, USA. Email: `jssmith@gatech.edu`

However, many social contexts require a more dynamic view of turn-taking roles. In collaborative scenarios, it is natural for one agent to take initiative for an extended period of time to make or suggest a contribution to common ground, during which time the interaction partner acts more passively. Then, roles may be switched. A completely dominant or completely passive setting of the CADENCE parameters would not allow all participants to make balanced contributions. Instead, the more natural behavior would be to switch dynamically between active and passive parameters.

In extending CADENCE to accommodate more dynamic turn-taking behavior, the question is then raised of when to switch roles. We hypothesize it is more socially appropriate to perform these changes at turn boundaries. We describe the relevant details of our original system and our extension to support dynamic role-switching, then present an experiment with 15 people interacting with that system in a multimodal object play task. Our results indicate that a robot can effectively change its timing parameters online to achieve a change in social dynamic for a human-robot dyad. This replicates the prior results, which were achieved from static parameter settings between groups, but in the more realistic interaction scenario of switching roles during a session. Our results also show that this change in social dynamic is only achieved if the robot makes its parameter change at a socially appropriate time and not at a fixed time interval.

II. Related Works

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 [2], [7]. Currently, it is common for interaction control to be modeled using a state-based representation such as a finite state machine (FSM) [11] or a partially observable Markov decision process (POMDP) [15], [12] 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 [13].

The coordination of joint action has been a topic of research in the domain of human-robot teamwork. Hoffman and Breazeal used Bratman's notion of shared cooperative activity [1] to develop a system that meshes human and robot subplans [5]. Shah developed a system for action coordination in HRI called Chaski [14], which focused on scheduling actions for both the human and the robot to minimize idle time. Other systems have also been developed to control multimodal dialogue for social robots, such as the work of [6] that controls dynamic switching of behaviors in the speech and gesture modalities, and the framework of [8] that controls task-based dialogue using parallelized processes with interruption handling. DiscoRT is dedicated towards real-time collaborative discourse using social modalities [9].

III. Timing Control in Human-Robot Dyads

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 [4]. Previously we have shown CADENCE can be parameterized to achieve different social dynamics [3]. In this paper we experiment with dynamically changing these parameters online. This section covers the key system components necessary for understanding our experiment.

A. Floor Regulation

Turn-taking involves the use of shared resources (e.g., speaking floor, shared space). For example, participants in an interaction can either hold the floor by speaking or audit the speaker. Good conversational turn-taking has a back and forth transfer of the floor; yielding by one party and seizing by the other. When this pattern is not followed, undesirable error state conditions in turn-taking are brought about, such as conflicts and lapses. Conflicts arise when both parties try to hold the floor at the same time. Lapses occur when both parties try to audit their partners at the same time. When humans and robots interact, the dynamics of the interaction are determined by the robot's turn-taking behavior and by the human's response to that behavior.

CADENCE is a context-free framework for turn-taking, allowing it to be applied in a wide range of domains. The goal is for domain specific behavior to be provided by a domain-specific context model. The context model determines what actions the robot will perform while CADENCE regulates when to execute them. The *Floor Regulator* is a component of CADENCE responsible for determining when the robot should take a turn, which is a set of modality-specific acts (including speech, gaze, gesture, and manipulation). The context model generates acts for each turn.

In order to regulate the robot's turn-taking behavior, CADENCE models the turn-taking states of both the human and the robot. Each can be in one of two turn-taking states: either holding the floor or auditing the interaction partner. The combination of these robot and human turn-taking states results in four possible *floor states* for the dyad: robot holding, human holding, conflict, and lapse. The floor is in a state of conflict when both human and robot attempt to take a turn simultaneously and is in a state of lapse when neither human nor robot attempts to take a turn. CADENCE models seizing as a transition from the auditing state to the holding state and yielding as a transition from the holding state to the auditing state. The amount of simultaneous holding or auditing required for conflict or lapse to be declared are the *conflict time* and *lapse time* parameters, respectively. The floor state is a function of the time the human and robot have spent in their current turn states.

Backchannels indicate intentions towards floor ownership. There are two types: *continuers* and *incipient speakership markers*. *Continuers* encourage the interaction partner to continue holding the floor while *incipient speakership markers* indicate the robot's desire to seize the floor. The specific acts for each include head gestures and spoken utterances. Backchannel actions are not interruptible and do not count towards the robot's floor time.

B. Seizing the Floor

At each time step while the robot is auditing its partner it can: take a full turn, backchannel, or do nothing. A primary factor in selecting which option is the floor time ratio: the ratio of the cumulative time the robot holds the floor to the cumulative time the person holds the floor. The robot attempts to steer the floor ratio towards a specified value, called the *floor factor*. When floor time ratio is less than *floor factor*, the robot's goal is to take a turn. If the person has yielded the floor, the robot may respond to its partner's previous turn after waiting for the *response delay* to pass. If the person continues holding, the robot can either backchannel an *incipient speakership marker* or interrupt the person. Interruptions are only possible if the *interrupt user* parameter is enabled and the person has held the floor for longer than the *interrupt patience*. When the floor time ratio is too high, the robot's goal is to avoid holding the floor. To this end, it may backchannel a *continuer* to encourage the person to continue holding the floor. Regardless of the floor time ratio, the robot will take a full turn if the dyad is in a state of lapse longer than the *lapse tolerance*. A shorter tolerance gives the person less time to seize.

C. Yielding the Floor

Once the robot has completed the actions in its turn it yields the floor and audits the partner. The robot may also yield the floor mid-turn if the person interrupts. This latter case only occurs if the *interrupt self* parameter is enabled. Such self interruptions are caused when the perception state transitions to *signaling* (i.e., the human is signaling a desire to hold the floor). This causes the robot to hesitate, pausing its actions and giving the person the chance to either back off or continue seizing the floor. This pause lasts for a period of time determined by the *hesitation resolution* parameter. If the perception state returns to *idle*, the robot resumes its turn. However, if the perception state transitions to *suppressing*, then the floor is in a conflict state. If this lasts longer than a period of time specified by the *conflict tolerance* parameter, the robot will abort its turn, yielding the floor.

D. Modeling the Human Partner

A key requirement for generating appropriate robot turn-taking behavior is modeling the human partner's floor state. We accomplish this by monitoring the person's physical movements and speech. Skeleton data from a Microsoft Kinect is used to track the locations of the person's joints over time. The person is classified as gesturing if her hands are currently inside the shared workspace in front of the robot or are in motion. The person's speech is captured using a microphone, and a Pure Data module [10] extracts the pitch of the signal. An audio sample is classified as speech when the pitch is greater than a specified threshold. The output values of the gesture and speech detectors are the fractions of samples over a sliding window for which their respective conditions have been satisfied.

From the output of these detectors, the person's behavior is classified as one of three perceptual states. *Idle* indicates low detector values: person is not performing actions. *Suppressing* indicates high detector values: behavior that attempts to suppress robot actions. *Signaling* indicates detector values that are between the previous two levels. From these, the person's floor state is classified as *holding* or *auditing*.

E. Turn-taking Parameters

Following is a summary of the CADENCE turn-taking parameters described above, for full details on the implementation described, the reader is referred to [3]:

- **Floor factor**: the desired ratio between the robot's and the human's holding of the floor
- **Response delay**: how long to wait after the human yields the floor before the robot can seize the floor
- **Interrupt user:** whether robot can interrupt human
- **Interrupt self:** whether human can interrupt robot
- **Conflict time:** how long robot and human spend in dual *holding* states before the floor state is a conflict
- **Lapse time:** how long robot and human spend in dual *auditing* states before the floor state is a lapse
- **Conflict tolerance:** how much conflict the robot will tolerate before yielding the floor
- **Lapse tolerance:** how much lapse the robot will tolerate before seizing the floor
- **Interrupt patience:** min amount of time person can continuously hold the floor before robot can interrupt
- **Hesitation resolution:** how long robot hesitates before resuming or interrupting its turn
- **Act spacing:** the minimum time between robot actions
- **Backchannel spacing:** min amount of time between backchannel actions

Our earlier work [3] studied people interacting with a humanoid robot that could speak, gesture, and manipulate objects. For each participant, the robot was run using either an active or a passive turn-taking parameterization. The passive parameterization resulted in longer lapses, yielding to user interruptions, and less floor holding, whereas the active parameterization resulted in the robot interrupting the user, ignoring user interruptions, and holding the floor more. The study drew three conclusions:

1) Changing the robot's turn-taking parameters did in fact result in different robot behavior, affecting overall interaction dynamics.
2) Different parameterizations resulted in different human behavior. People who interacted with the robot in a passive parameterization talked more than those who interacted with the active parameterization.
3) Participants attributed different personality types to the robot based on which parameterization they observed.

F. Changing Social Dynamics

These effects were all demonstrated between different groups of people interacting with the robot. This leads to the focus of the current work: enabling the ability to change between parameterizations during a single interaction.

We extended CADENCE to allow it to intelligently switch between parameterizations based on the state of the turn-taking interaction. In order to maintain consistent robot behavior during turns, we hypothesized that parameter changes should only occur on turn boundaries, marked by transitions to *seize* and to *yield*. Changing from passive to active parameters while the person is taking a turn could cause the robot to immediately interrupt the person, while changing from active to passive parameters during the robot's turn could cause the robot to interrupt its own turn. Changing from active to passive parameters when *seizing* the floor could cause similar problems. If the person continues talking after the robot interrupts, the switch to passive parameters could force the robot to interrupt itself to avoid conflict.

Consequently, we chose to switch from passive parameters to active parameters only when *seizing* the floor and from active parameters to passive parameters only when *yielding* the floor. Floor times are cleared every time the turn-taking parameters are changed to ensure that the ensuing robot behavior is independent of past parameterizations.

IV. Experiment

In this experiment, we investigate whether behavioral differences are exhibited during a human-robot interaction when turn-taking parameters are changed during the interaction. We also investigate whether turn boundaries are more appropriate inflection points for changing social dynamics when compared to a simple time-based strategy.

A. Hardware Platform

We implemented our turn-taking system on a Meka Robotics M3 humanoid robot, named Curi. It has 7-DOF arms and 4-DOF hands permitting object manipulation and gestures. It also has an expressive head and neck, allowing gaze and head gestures. The robot is situated in front of a kitchen counter that is located between the robot and

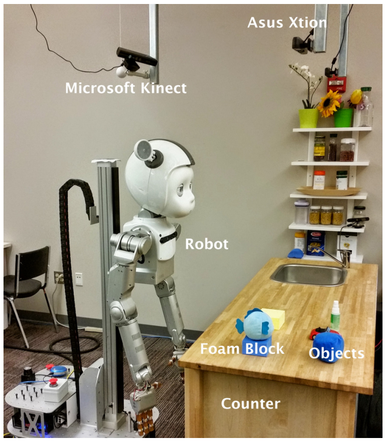

Fig. 1. The experimental setup

the person (see Fig. 1). An Asus Xtion RGB-D sensor mounted overhead provides data regarding objects placed on the countertop, while a Microsoft Kinect located behind the robot and a wireless microphone worn by the person provide data on the participant's state.

B. Task Domain

This experiment uses an open-ended domain, similar to [3]. In this domain, the robot and the human talk about, gesture at, and manipulate objects on a table. To reduce the effects of domain semantics on turn-taking behavior, the robot uses an artificial language. The open-ended nature of the domain is intended to capture the balanced give and take of child's play, in which both participants can comment, narrate, express attitudes, or teach each other about objects. The context model for this domain provides actions for several modalities:

- **Speech**: Robot speech is generated from random sequences of phonemes. Phrases for full turns use a different set of phonemes than phrases for backchannels. The use of an artificial language forces the person to make turn-taking decisions based on the timing of actions rather than context-specific cues.
- **Gaze:** The robot can gaze at salient locations, such as objects on the table or the participant's face and hands.
- **Object manipulation:** The robot can pick and place or point to objects on the table. Manipulation acts are only performed on light objects of a specific size placed on foam blocks located at predetermined locations.
- **Arm gestures:** The robot can perform several prerecorded gestures with its arms. These gestures are roughly equivalent to a human shrugging and may be performed with or without an object in hand.

The turn-taking parameters defined by the context model are as indicated in Table I.

C. Experiment Design

We compare two different conditions that trigger parameter changes: the Time Condition and the Floor Condition.

TABLE I
PARAMETERS THAT DIFFERED BETWEEN TWO CONTRASTING TURN-TAKING ROLES

Parameter	Active	Passive
Floor factor	2.0	0.5
Interrupt user	true	false
Interrupt self	false	true
Conflict tolerance	N/A	1000 ms
Lapse tolerance	500 ms	4000 ms
Act spacing	50–250 ms	500–1000 ms
Backchannel spacing	1000–2000 ms	3000–4000 ms

The Floor Condition is intended to switch robot behavior at natural turn-taking junctures in the interaction. Since the domain is free of semantic constraints on such junctures, a countdown timer is used to schedule parameter changes. Once this timer reaches zero, the Floor Condition switches parameters according to the turn boundary constraints described in Sec. III-F.

As a baseline for comparison, the Time Condition changes parameters as soon as the countdown timer reaches zero.

Our experimental model leads us to these hypotheses:

- **Hypothesis 1: People will speak more during the periods when the robot uses passive parameters than during the periods with active parameters.** Our previous work [3] found that people who interacted with the robot in an exclusively passive parameterization talked more than those who interacted with the robot in an exclusively active parameterization. We speculate that similar differences in behavior will be seen within an interaction if the robot switches back and forth between the two parameterizations.
- **Hypothesis 2: The Floor Condition will lead to more significant changes in participant speaking behavior than the Time Condition.** By delaying turn-taking behavior changes to turn boundaries, the Floor Condition should minimize inconsistent behavior mid-turn, thereby facilitating people's ability to successfully adapt to the behavior change that the robot initiates.

D. Experiment Protocol

The participants were 15 undergraduate student volunteers, 10 male and 5 female. They were randomly assigned to either the Floor Condition or Time Condition. Each person interacted with the robot for two sessions, approximately three minutes each. Whether they started with an active or passive parameterization was counter-balanced across the groups. During each session, parameters were switched three times, resulting in four segments per session. The Time Condition segments were exactly 45 seconds long. Floor Condition segments were at least 45 seconds long but depended on the behavior of the person. Since the Floor Condition only loads active parameters when the robot seizes the floor, a person can prolong a passive segment by talking continuously.

Condition	Passive	Active	p
Time	0.52 (0.12)	0.47 (0.18)	0.59
Floor	0.42 (0.11)	0.28 (0.12)	**0.02**
p	0.14	**0.02**	

TABLE II

NORMALIZED HUMAN SPEAKING TIMES

Condition	Passive–Active	Active–Passive	p
Time	0.03 (0.17)	0.0 (0.11)	0.71
Floor	0.15 (0.08)	-0.18 (0.13)	$<$**0.001**
p	0.17	**0.03**	

TABLE III

CHANGES IN NORMALIZED HUMAN SPEAKING TIMES BETWEEN ADJACENT SEGMENTS OF DIFFERENT PARAMETER SETTINGS

The participants were instructed to teach the robot about a set of objects on the countertop. They were told the robot would behave like a child, still too young to speak properly, so they would not understand the robot's speech. Participants were not informed of the different parameterizations or conditions, nor were they told that the robot could only detect the presence or absence of speech. They were asked to rest their arms at their sides rather than on the countertop to prevent the robot's Kinect tracking from falsely interpreting an attempt to seize the floor.

E. Measures and Analysis

We assess the behavior of people during each session by two metrics: amount of time spent speaking and amount of conflict. Some conflict (simultaneous speaking by both parties) normally occurs during social interactions. However, conflict is a useful metric since greater conflict correlates with poorer turn-taking performance.

In order to compare people's behavior between different robot turn-taking parameterizations, we divide each interaction into segments at the junctures of parameter changes. Each segment corresponds to a time period when the robot exhibits only passive or only active turn-taking behavior. Speech and conflict times are summed over each segment. To compensate for the varied durations of segments in the Floor Condition, the values of these metrics within each segment are normalized by dividing them by the segment's duration.

V. RESULTS

We originally planned to analyze both the first and second sessions. However, the disparity in first session lengths between the Time Condition and Floor Condition was greater than anticipated.[1] After removing all participants whose first session was compromised,[2] we were left with 8 participants, two per combination of condition and initial robot parameterization. Half of these remaining participants were female, representing 3 out of 4 sessions for the Time Condition and 1 out of 4 sessions for the Floor Condition.

[1]Since the Floor Condition postpones a switch to active parameters until the person finished their turn, their first sessions were at least 30-60 seconds longer than people in the Time Condition. Any analysis from the second sessions of the two conditions could simply be a result of this disparity.

[2]Some technical issues forced us to discard some users' sessions. The most severe muted the robot's speech for portions of several participants' sessions, preventing them from knowing when the robot was speaking.

Condition	Passive	Active	p
Time	0.16 (0.08)	0.33 (0.15)	**0.01**
Floor	0.16 (0.09)	0.17 (0.11)	0.83
p	0.99	**0.03**	

TABLE IV

NORMALIZED SPEAKING CONFLICT TIMES

A. Participant Behavior between Parameterizations

To determine whether participants displayed different behavior during the two different turn-taking parameterizations, we compared the normalized amounts of time that participants spent speaking during the active segments to the normalized amounts of time they spent speaking during the passive segments (Table II). Participants in the Floor Condition spoke significantly more during passive segments (M = 0.42, SD = 0.11) than during active segments (M = 0.28, SD = 0.12) (p = 0.02). We did not find a significant difference between parameterizations in the normalized amount of time spoken by participants in the Time Condition. Participants adapted more readily to the socially appropriate behavioral changes of the Floor Condition than to the arbitrary changes of the Time Condition.

To see whether alternating robot turn-taking parameterizations results in an alternating pattern of participant behavior, we looked at the difference in participant speech between *adjacent* segments (Table III). For the Floor Condition, there was a significant difference between switching from Active to Passive segments (M = 0.15, SD = 0.08) and from Passive to Active segments (M = -0.18, SD = 0.13) (p $<$0.001). Participants spoke significantly more when the robot switched to a passive parameterization and significantly less when it switched to an active parameterization. We did not see a significant difference for the Time Condition.

B. Turn-Taking Quality between Conditions

In order to evaluate the quality of turn-taking, we analyzed conflict during the different conditions and parameterizations (Table IV). An increase in conflict means that more of the interaction was spent with both parties attempting to hold the floor simultaneously, suggesting a breakdown in agreement over whose turn it is. During the robot's passive parameterization, there was no significant difference in conflict between the Time Condition and the Floor Condition. The passive parameters limit conflict by taking less floor time and directing the robot to quickly interrupt itself if conflict occurs. During the active parameterization, the robot attempts to hold the floor longer and does not yield to interruptions by people. This metric provides an indication as to how well people respected the robot's turn initiative. During the robot's active parameterization, conflict was significantly greater in the Time Condition (M = 0.33, SD = 0.15) than in the Floor Condition (M = 0.17, SD = 0.11) (p = 0.03). This suggests the robot is more successful at holding the floor in the Floor Condition than in the Time Condition.

VI. Discussion and Limitations

Our results supported both of our experimental hypotheses from Section IV-C. Hypothesis 1 was supported in the Floor Condition, in which people spoke significantly more when the robot was passive and significantly less when the robot was active. Hypothesis 2 was also supported by our data; humans did not change their speaking behavior significantly in the Time Condition in response to robot parameter changes, but did in the Floor Condition. The more appropriate robot behavior in the Floor Condition also led to fewer speaking conflicts.

We found it surprising that the effect of robot parameter changes was so minor in the Time Condition. While passive and active settings resulted in contrasting conflict times, they did not tend to impact the human's speaking behavior. Just three badly misplaced switches can confuse humans for an entire interaction about when they should or should not speak, nullifying any mutual adaptivity to be gained from switching roles. These results highlight the importance of each individual turn decision in an interaction.

One limitation of our analysis is that we did not include lapse times as a metric for turn-taking breakdowns. Speech was the only modality that could be automatically extracted with reliability, and lapses are not uniquely determined by absence of speech from both parties. For example, hand gestures or manipulation in the absence of speech would not constitute a lapse. In contrast, extended simultaneous speech is almost always interpreted as conflict. Because our perception system could not detect human actions reliably, accurate lapse analysis requires human annotation.

Another limitation of the experiment was that conditions were not perfectly gender balanced. After removing contaminated sessions, the Time Condition was majority female and the Floor Condition was majority male. This may have affected our results.

VII. Conclusion and Future Work

We previously demonstrated that using contrasting static robot turn-taking parameterizations resulted in measurable behavioural differences in both the human and the robot. This paper demonstrates that human behavior can also be significantly affected by alternating the different turn-taking parameters within a single human-robot interaction. Through an experiment with 15 participants, we find that people interacting with a humanoid robot speak more when the robot uses passive parameters and less when the robot uses active ones. We also find that people alter their behavior more when the robot is programmed to switch turn-taking behavior at natural interaction points (represented by the Floor Condition) rather than in response to a fixed countdown timer (represented by the Time Condition).

These findings motivate further research into varying robot behavior to promote a desired effect in human behavior or interaction outcome. Since these behavior changes were performed in a generic context, it is likely they could apply to a wide variety of domains. For future work, we intend to utilize this parameter-switching system as part of establishing leader/follower roles for a robot in a collaborative setting.

References

[1] M. Bratman. Shared cooperative activity. *The Philosophical Review*, 101(2):327–341, 1992.

[2] J. Cassell, T. Bickmore, L. Campbell, K. Chang, H. Vilhjálmsson, and H. Yan. Requirements for an architecture for embodied conversational characters. In *Proceedings of Computer Animation and Simulation*, pages 109–120, http://dx.doi.org/10.1007/978-3-7091-6423-5_11, 1999.

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

[4] C. Chao and A. L. Thomaz. Timed petri nets for multimodal interaction modeling. In *Proceedings of the ICMI 2012 Workshop on Speech and Gesture Production in Virtually and Physically Embodied Conversational Agents*, 2012.

[5] G. Hoffman and C. Breazeal. Collaboration in human-robot teams. In *Proceedings of the 1st AIAA Intelligent Systems Conference*, 2004.

[6] T. Kanda, H. Ishiguro, M. Imai, and T. Ono. Development and evaluation of interactive humanoid robots. In *Proceedings of the IEEE*, volume 92, pages 1839–1850, http://dx.doi.org/10.1109/JPROC.2004.835359, 2004.

[7] B. Mutlu, T. Shiwa, T. K. H. Ishiguro, and N. Hagita. Footing in human-robot conversations: How robots might shape participant roles using gaze cues. In *Proceedings of the 2009 ACM/IEEE Conference on Human-Robot Interaction (HRI)*, pages 61–68, http://dx.doi.org/10.1145/1514095.1514109, 2009.

[8] M. Nakano, Y. Hasegawa, K. Nakadai, T. Nakamura, J. Takeuchi, T. Torii, H. Tsujino, N. Kanda, and H. Okuno. A two-layer model for behavior and dialogue planning in conversational service robots. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3329–3335, 2005.

[9] B. Nooraei, C. Rich, and C. L. Sidner. A real-time architecture for embodied conversational agents: Beyond turn-taking. In *Proceedings of the 7th International Conference on Advances in Computer-Human Interactions (ACHI)*, 2014.

[10] M. Puckette. Pure Data: another integrated computer music environment. In *Proceedings of the International Computer Music Conference*, pages 37–41, 1996.

[11] A. Raux and M. Eskenazi. Optimizing the turn-taking behavior of task-oriented spoken dialog systems. *ACM Transactions on Speech and Language Processing*, 9(1):1–23, http://doi.acm.org/10.1145/2168748.2168749, 2012.

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

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

[14] 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 (HRI)*, pages 29–36, 2011.

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