Anticipation in Human-Robot Interaction

Guy Hoffman
Georgia Tech Center for Music Technology
840 McMillan St
Atlanta, Georgia 30332

Abstract

Anticipating the actions of others is key to coordinating joint activities. We propose the notion of anticipatory action and perception for for robots acting with humans. We describe four systems in which anticipation has been modeled for human-robot interaction; two in a teamwork setting, and two in a human-robot joint performance setting. In evaluating the effects of anticipatory agent activity, we find in one study that anticipation aids in team efficiency, as well as in the perceived commitment of the robot to the team and its contribution to the team's fluency and success. In another study we see anticipatory action and perception affect the human partner's sense of team fluency, the team's improvement over time, the robots contribution to the efficiency and fluency, the robot's intelligence, and the robots adaptation to the task. We also find that subjects working with the anticipatory robot attribute more human qualities to the robot, such as gender and intelligence.

Introduction

In order to design robots that can work fluently with human partners in a physically situated setting, we want to overcome the delayed turn-taking structure of most human-robot interaction. Piecewise interaction often follows from the stepwise perception-action paradigm, and the discretized modeling of decision making, which is at the base of most interactive reasoning and dialog systems in AI literature. Contrast that to the simultaneous "dance" of two human agents performing together at high level of coordination and adaptation, in particular when they practice a task repetitively, and are well-accustomed to the task and to each other.

In recent years, the cognitive mechanisms of joint action have received increasing attention. Among other factors, successful coordinated action has been linked to the formation of expectations of each partner's actions by the other and the subsequent acting on these expectations (Sebanz, Bekkering, and Knoblich 2006; Wilson and Knoblich 2005). We argue that the same holds for collaborative robots: if they are to go beyond stop-and-go interaction, agents must take into account not only past events and current perceived state, but also an anticipatory model of their human collaborators, and apply anticipatory action as a result of these models. We therefore propose that the timing and meshing of anticipatory action and perception are a useful framework for human-robot interaction.

In this paper, we describe our work over the last few years in anticipatory action for human-robot interaction. We detail four systems in which we have modeled anticipation for agents acting jointly with humans. In two systems, an agent collaborates on a team task with a human team member. We propose two models of anticipatory action and evaluate their effect on human subjects. We apply a slightly different anticipatory approach in two additional systems, aimed at human-robot stage performance. One is a theater performance robot, and another a musical improvisation robot. Both systems have been successfully used in a live on-stage performance with human actors and musicians, in front of live audiences. In all four cases, our aim was to design systems that not only achieve the task at hand, but do so in a highly coordinated, fluent fashion.

Related Work

Most work related to joint action has been concerned with a goal-oriented view of the problem, paying little attention to the timing of actions, or the quality of action meshing. Joint action is usually described as solving a problem where the participants share a goal and a common plan of execution.

In Bratman's analysis of Shared Cooperative Activity, for example, he defines certain prerequisites for an activity to be considered shared and cooperative (Bratman 1992), such as mutual responsiveness, commitment to the joint activity, and commitment to mutual support. Supporting Bratman's guidelines, Cohen and Levesque propose a formal approach to building artificial collaborative agents (Levesque, Cohen, and Nunes 1990). Their notion of joint intention is viewed not only as a persistent commitment of the team to a shared goal, but also implies a commitment on part of all its members to a mutual belief about the state of the goal. These and similar principles have been used in a number of human-robot teamwork architectures (Hoffman and Breazeal 2004; Alami et al. 2005).

Human-robot collaboration has been investigated in a number of previous works, although the question of timing or fluent action meshing has not received much attention. (Kimura, Horiuchi, and Ikeuchi 1999) have studied a robotic

arm assisting a human in an assembly task. Their work addressed issues of vision and task representation, but does not deal with anticipation or timing. Other human-robot collaboration work, such as that of (Jones and Rock 2002) studies human-robot collaboration with an emphasis on dialog and control, aimed primarily at the teleoperation scenario.

Some work in shared-location human-robot collaboration has been concerned with the mechanical coordination and safety considerations of robots in shared tasks with humans, e.g. (Khatib et al. 2004) . Other work addresses turn-taking and joint plans, but not anticipatory action, practice, or fluency (Hoffman and Breazeal 2004). Timing and synchronization have been reviewed on the motor level in the context of a human-robot synchronized tapping problem (Komatsu and Miyake 2004). Anticipatory action, without relation to a human collaborator has been investigated in the area of robot navigation, e.g. (Endo 2005).

Anticipatory Action in a Non-Atomic MDP

Our first anticipatory action system is built around a time-aware extension of a Markov Decision Process (MDP). MDPs are useful structures to model decision making under uncertainty. However, they model perception and action as atomic entities enforcing a turn-based activity paradigm which may be inappropriate to time-based joint activities between agents and humans.

In order to incorporate time into the decision process, we have proposed an extension of MDPs in which two agents share a common workspace, on which the actions of both have effect, and in which action is non-atomic.

In our model, the two agents have a number of internal states, and the workspace has a number of external states, which the agents can both perceive and affect. Human and robot have distinct action sets, and a transition function maps state-action pairs to new states.

To allow us to investigate temporal aspects of the actions of two collaborating agents, state transitions are not atomic, and the decision to take a particular action does not result in an immediate state transition. Instead, moving between states takes time, and is associated with a known discrete cost, which is a function of the states before and after the action. This cost can be thought of as the 'distance' between states, or more generally — the duration it takes to transition between states.

The Factory World

In our experiments we use a simulated factory setting (Figure 1). The goal of the team is to assemble a cart made of different parts. The labor is divided between the human and the robot: the human has access to the individual parts, and is capable of carrying them and positioning them on the workbench. The robot is responsible for fetching the correct tool and applying it to the currently pertinent component configuration in the workbench. Each tool has to be returned to its stock location. The size of the state-space in this simulation is 2,160,900.

The action-space of the robot includes mobility actions, moving to one of the five locations in the factory. In addition

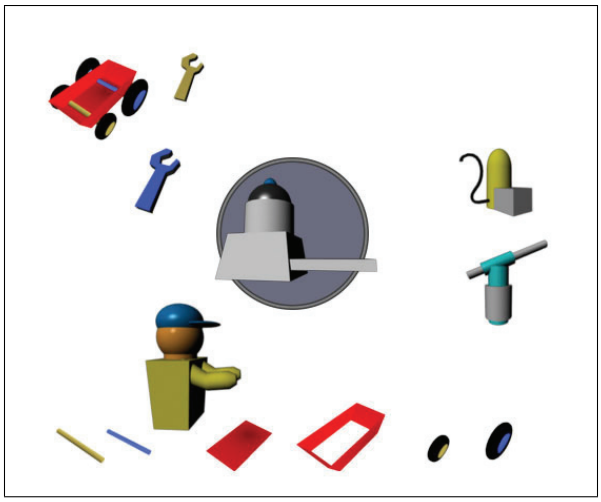

Figure 1: Simulated factory setting with a human and a robot building carts, while sharing a workbench, but dividing their tasks.

the robot can pick up and put down a tool, as well as use it on the workspace.

Reactive vs Anticipatory Action

We compare two kinds of artificial agents operating in this framework: reactive and anticipatory agents. Reactive agents apply their decision function on the currently perceived state. A number of strategies are possible for this agent class, but they share the trait that they only take into account the current state.

In contrast, the agent can take an anticipatory, "risk-taking" policy, in which it acts on a combination of the existing state and a probabilistic view of the temporal activity of the human teammate. In the factory world described above, we model the probabilistic prediction of the human's actions as a first-order Markov Process. The agent learns the parameters of this Markov process using a naïve Bayesian estimate. Using its knowledge of action durations, the agent can then estimate the expected temporal cost of each action based on the probabilistic model of the human's action. The anticipatory selects actions based on an expectation-optimization strategy, using a risk-aversion parameter.

Analysis

We showed that using an anticipatory expected-cost minimizing strategy for action selection results in a theoretical improvement in efficiency over the reactive strategy. This improvement in efficiency increases over time as the model of the human behavior gets reinforced, and is dependent on the human's consistency.

We also tested our system in a human subject study involving untrained human team members. One group collaborated with a reactive agent, and one with an anticipatory agent. In the post-experimental survey, we found significant differences between participants in the two groups. Subjects

in the anticipatory action agent group rated the robot as significantly higher when asked whether "The robot's performance was an important contribution to the success of the team."; "The robot contributed to the fluency of the interaction."; and even in the emotionally charged "it felt like the robot was committed to the success of the team."

We also found a significant difference in the percentage of concurrent motion, and perceived delay between the human's actions and the robot's actions, as well as anecdotal evidence of a more positive attitude towards the anticipatory agent in open-ended questions. For a full analysis of this system, please refer to (Hoffman and Breazeal 2007).

Anticipatory Action as Perceptual Simulation

In our second implementation, we extend the above-described system to cover real-world perceptual input and continuous decision-making, instead of stepwise probabilistic evaluation of expected cost, and simulated perception.

Neuropsychological analysis of anticipation in human joint activities and teamwork points towards a perceptual simulation framework (Wilson and Knoblich 2005; Sebanz, Bekkering, and Knoblich 2006), according to which agents simulate perceptual activation in anticipation of external events and the actions of their collaborators.

Based on these findings, we propose that anticipation through perceptual simulation can provide a powerful model for robots acting jointly with humans if they are to collaborate fluently using multi-modal sensor data. To that end, we developed a cognitive architecture based on the principles of anticipatory top-down perceptual simulation.

Cognitive Model

Our approach posits that an improvement in the timing of joint actions achieved through repetitive practice can be achieved through a system that relies on two processes: *(a) anticipation* based on a model of repetitive past events, and *(b)* the modeling of the resulting anticipatory expectation as *perceptual simulation*, affecting a top-down bias of perceptual processes.

In this model, perceptions are processed in modality streams built of interconnected process nodes. These nodes can correspond to raw sensory input (such as a visual frame or a joint sensor), to a feature (such as the dominant color or orientation of a sensory data point), to a property (such as the speed of an object), or to a higher-level concept describing a statistical congruency of features, in the spirit of the Convergence Zones in (Simmons and Barsalou 2003).

Modality streams are connected to an action network consisting of action nodes, which are activated in a similar manner as perceptual process nodes. An action node, in turn, leads to the performance of a motor action. Connections between nodes in a stream are not binary, but weighted according to the relative influence they exert on each other.

Importantly, activation flows in both directions, the *afferent*—from the sensory system to concepts and actions—and the opposite, *efferent*, direction. If an activation is triggered in the efferent pathway, the result is a simulation of a "priming"-like phenomenon, as follows:

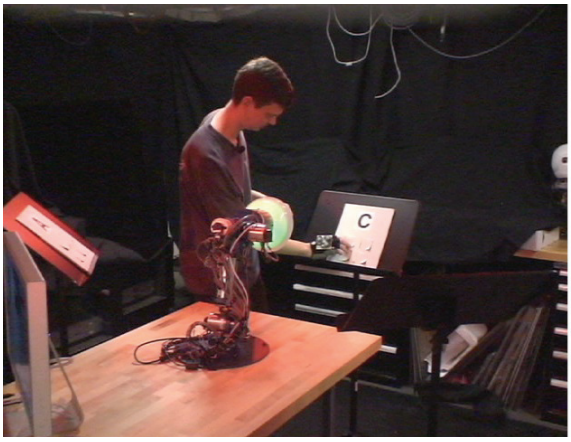

Figure 2: The collaborative lighting task workspace.

If a certain higher-level node is activated, the lower-level nodes that feed that higher-level node are partially activated through simulation on the efferent pathway. As this activation is added to the sensory-based activation in the lower-level nodes, this top-down activation inherently lowers the perceptual activation necessary for the activation of those lower-level nodes, decreasing the real-world sensory-based activation threshold for action triggering. The result of this is reduced response time for anticipated sensory events, and increasingly automatic motor behavior for rehearsed perception-action congruences.

Two subsystems support anticipation in this framework: the first is a Markov-chain Bayesian predictor, building a probabilistic map of node activation based on recurring activation sequences during practice, similar to the one described in the previous Section. It triggers high-level simulation, which—through the modality stream's efferent pathways—biases the activation of lower-level perceptual nodes. If the subsequent sensory data supports these perceptual expectations, the robot's reaction times are shortened as described above. In the case where the sensory data does not support the simulated perception, reaction time is longer and can, in some cases, lead to a short erroneous action, which is then corrected by the real-world sensory data. While slower, we believe that this "double-take" behavior, often mirrored by the human partner's motion, may contribute to the human's sense of similarity and bond to the robot.

An additional mechanism of practice is that of weight reinforcement on existing activation connections. While most node connections are fixed, some can be assigned to a connection reinforcement system, which will dynamically change the connection weights between the nodes. This system works according to the contingency principle, reinforcing connections that co-occur frequently and consistently, and decreasing the weight of connections that are infrequent or inconsistent. This subsystem thus reinforces consistent coincidental activations, but inhibits competing reinforcements stemming from the same source node, leading to anticipated simulated perception of inter-modal perception nodes. This, again, triggers top-down biasing of lower-level

perception nodes, shortening reaction times.

Application

We have implemented an instantiation of the proposed architecture on two robots: a humanoid robot, and a robotic desk lamp, depicted in Figure 2. The lamp has a 5-degree-of-freedom arm and a LED lamp which can illuminate in the red-green-blue color space. The robot employs a three-modality sensory and perceptual apparatus, including a 3D vision system, audio with speech recognition, and a proprioceptive perceptual stream fed from the robot's motor controllers.

In the human-robot collaboration used in our studies, the human operates in a workspace as depicted in Figure 2. The robot can direct its head to different locations, and change the color of its light beam. When asked to "Go", "Come", or "Come here" the robot would point towards the location of the person's hand. Additionally, the color changed in response to speech commands to one of three colors: blue, red, and green. The workspace contained three locations (A,B,C). At each location there was a white cardboard square labeled with the location letter, and four doors cut into its surface. Each door, when opened, revealed the name of a color printed underneath. The task was to complete a sequence of 8 actions, which was described in diagrammatical form on a sequence sheet, and included opening a door and shining the right-colored light onto the carboard. This sequence was to be repeated 10 times, as quickly as possible.

Analysis

In a human subject study involving untrained human team mates, we have confirmed a number of behavioral hypotheses relating to the efficiency and fluency of the human-robot team. Again, subjects were divided into two groups, `REACTIVE` (bottom-up only) and `ANTICIPATORY` (top-down), and showed a significant difference on overall task time, mean sequence time, human idle time, and perceived delay (Hoffman and Breazeal 2008b).

In a questionnaire about the subjects' experience with the two robots, we find significant differences between subjects in the two experimental conditions: subjects in the `ANTICIPATORY` condition perceived the robot's behavior as more fluent, and rated the robot's contribution to the team, as well as the team's overall improvement, significantly higher than subjects in the `REACTIVE` condition. This supports our hypothesis that the proposed architecture contributes to the quality of timing and collaboration in human-robot teams.

In their open-ended responses, a lexical analysis shows that subjects in the `ANTICIPATORY` condition commented on the robot more positively, and subjects in the `REACTIVE` condition commented on the robot more negatively. `ANTICIPATORY` subjects attributed more human characteristics to the robot, although there is little difference in the emotional content of the comments. Also, gender attributions, as well as attributions of intelligence occurred only in the `ANTICIPATORY` condition, while subjects in the `REACTIVE` conditions tended to comment on the robot as being unintelligent. We also found self-deprecating comments more prevalent in the `ANTICIPATORY` condition (Hoffman and Breazeal 2008a).

Anticipation in Human-Robot Theater

An additional class of joint activities, in which timing plays an important role, is that of a live human-robot stage performance. We have explored a different notion of anticipation for synchronized action meshing between humans and robots on stage, both in a theatrical and a musical setting.

The challenge of designing a system to control a live robot interacting on stage with human actors is to enable the robot to be both expressive and responsive. Most existing systems fall on one extreme of the scripted/direct-drive spectrum: they either are triggered in real time, but do not allow for a continuous expressive, or precisely animated, performance, or they are expressively animated but do not allow for precisely-timed reactive behavior.

To straddle this gap, we have developed a hybrid control system aimed for rehearsal and production of live stage performances of robots acting with humans. This system is intended to allow a single operator to control a robotic actor using pre-animated gestures and sequences, but at the same time adapting to the rhythm of live performance to the human actors. The result permits the robot to be both expressive and responsive to its scene partner.

Cue-Impulse Separation

Our solution lies in the possibility to trigger a scene action through a combination of anticipatory and follow-through movements. The base narrative layer is structured around the play's scenes. A scene is a sequence of beats, each of which describes a gesture on the robotic character's part.

To allow for complex gesture expressiveness, a scene is animated in a 3D animation software, using a physically structured model of the robot. This results in a sequence of positions for the robot throughout the scene, broken into frames. A custom-written exporter to the animation program exports the robot's DoF positions in radians for each of the frames in the scene, which are saved in the scene database.

Next, beats are identified and delimited in each scene. A beat is defined by an onset frame and end frame. During performance, a beat is expressed in two parts: the anticipatory *impulse* and the follow-through *cue*, two terms borrowed from acting method. To quote acting guru Sanford Meisner: "[T]he impulse comes early in the speech, and the cue then plays that out." (Meisner and Longwell 1987) The beat's impulse is the preparatory behavior of the character, which happens before the character's cue to perform an action, as an initial reaction to the scene partner's action. In order to support this in our system, a beat is assigned two speeds, in frames per second, for the impulse-to-cue, and cue-to-end parts of the beat. The result of this impulse-to-cue architecture is to prevent a stop-and-go delayed performance on the robot's part, and allowing for a fluent exchange of movement on stage.

Performance

We staged a theater production using the above-mentioned system in three live performances in front of an audience

Figure 3: Scene from a stage production employing the described hybrid puppeteering system.

of roughly 50 each night. The performed play was entitled *Talking to Vegetables*. The robotic performer used in the play was the same robotic desk lamp described in the previous Section. Figures 3 show a photo of the robot in the production. For a full description of the system, the performance, and responses to the show, please refer to (Hoffman, Kubat, and Breazeal 2008).

Anticipation in Human-Robot Musicianship

In a similar vein, we have applied an anticipation/followthrough framework to a human-robot joint musical performance. This work is as part of our work in robotic musicianship (Weinberg and Driscoll 2006). Few applications are as sensitive to time as music, and in particular the case of a joint musical performance. We have tried to build a system that can perform live with a human performer, improvising freely, and without noticeable delay.

Our system was implemented on Shimon, a robot playing a percussion instrument called marimba. The physical robot is comprised of four arms, each actuated by a voice-coil linear actuator at its base, and running along a shared rail, in parallel to the marimba's long side. The arms are custom-made aluminum shells housing two rotational solenoids each. The solenoids control mallets, chosen with an appropriate softness to fit the area of the marimba that they are most likely to hit. Each arm contains one mallet for the bottom-row ("white") keys, and one for the top-row ("black") keys. *Shimon* was designed in collaboration with Roberto Aimi of *Alium Labs*.

Anticipatory Action

We describe a musical improvisation system in a separate paper (in review). The system includes a number of interaction modules, such as call-and-response, joint improvisation, and accompaniment. In all of these modules, it is crucial for the robot to be able to strike the correct notes at the precise time, and often together with the cue from the human player.

Since it takes time for the robot to reach the appropriate position to play the correct note, we achieve synchronization by taking an anticipatory action approach, dividing gestures into *preparation* and *follow-through*, similar to the robotic theater control system. By separating the—potentially lengthy—preparatory movement (in our case: the horizontal movement) from the almost instant follow-through (in our case: the mallet action), we can achieve a high level of synchronization and beat keeping without relying on a complete-musical-bar delay of the system, as usually done in real-time joint music systems.

For example, in the call-and-response module, the robot prepares the response chord while the human is playing their phrase. If the phrase matched the prepared chord, the robot can respond with the first note of the response almost instantly. In the joint improvisation module, the robot starts the anticipatory action between the last beat of a bar and first beat of the next bar. This enables it to follow through at the precise onset of each bar. Similarly, for the grand finale of a performance, the robot's anticipatory movement is triggered by the human's crescendo, with the final chord activated by the human's cue, resulting in a seemingly synchronized musical experience.

Performance

We have used the described interaction module as part of a live human-robot Jazz performance before a public audience in April 2009 in Atlanta, GA, USA. The performance was part of an evening of computer music and was sold-out to an audience of approximately 160 attendants.

Figure 4: Live performance of the robot $Shimon$ using the improvisation system discussed herein.

The performance was structured around a "Jordu", a Jazz standard by Duke Jordan. The overall performance lasted just under seven minutes. Video recordings of the performance were widely covered by the press and viewed by an additional audience of over 40,000 online viewers.

Visual contact and Synchronization

In experiments with experienced pianists, we evaluated the role of visual contact using the anticipatory gestures described above. In particular, we used the call-and-response module from the above-mentioned performance system, in three conditions: in one, the robot was in plain sight of the pianist; in the second the robot was physically present, but occluded; and in the third, the music was synthesized by a computer without any physical movement of the robot.

We find that, when the robot slightly alters the tempo in response to the human's playing, the pianists' ability to synchronize with the robot is significantly reduced, compared

to the robot playing precisely on tempo. In these case, visual contact significantly reduces the error compared to the occluded and synthesized condition. In particular, visual contact allows the pianists to *react* to the robot instead of pre-empting the timing of their playing (often badly). This indicates that the pianists use the visual cues to time their playing, making use of the robot's anticipatory gestures to time their own musical activities.

We also find that visual contact is more crucial during slow trials, and during trials in which the robot slows down, possibly suggesting that visual cues are slow to be processed and aid less in fast sequences. It may be that during fast sequences, the pianists did not have time to look at the robot. We report on these and other results from this study in a separate paper (in review).

Conclusion

In this paper, we survey a number of projects exploring the notion of anticipation in human-robot interaction. We present a time-based extension to a Markov Decision Process, in which we developed an anticipatory agent to work together with a human in a joint state space. Comparing the anticipatory agent to a purely reactive agent, we find that human subjects collaborating with the agents find the anticipatory agent to be more fluent and more committed.

We also model anticipation as perceptual simulation, based on neurological findings in humans. We show a system using top-down simulation to achieve perceptual anticipatory action, and discuss an implementation on a collaborative robotic lamp. We present a human subject study evaluating the effects of our approach, comparing it with a system using only bottom-up processing. We find significant differences in the task efficiency and fluency between the two conditions. From self-report, we find significant differences in the perception of the team's fluency and the robot's contribution to that fluency, as well as in a number of other self-report metrics. Interestingly, we also find a tendency towards self-criticism in subjects collaborating with the anticipatory version of the robot.

Finally, we present two performance systems, one in the realm of theater performance, and one in the musical field, in which a apply an anticipation/followtrough framework to achieve a high level of synchronization with human stage partners. In the musical application, we also find that human players use visual contact to synchronize with a robot using the robot's anticipatory movement.

References

Alami, R.; Clodic, A.; Montreuil, V.; Sisbot, E. A.; and Chatila, R. 2005. Task planning for human-robot interaction. In *sOc-EUSAI '05: Proceedings of the 2005 joint conference on Smart objects and ambient intelligence*, 81–85. New York, NY, USA: ACM Press.

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

Endo, Y. 2005. Anticipatory and improvisational robot via recollection and exploitation of episodic memories. In *Proceedings of the AAAI Fall Symposium*.

Hoffman, G., and Breazeal, C. 2004. Collaboration in human-robot teams. In *Proc. of the AIAA 1st Intelligent Systems Technical Conference*. Chicago, IL, USA: AIAA.

Hoffman, G., and Breazeal, C. 2007. Cost-based anticipatory action-selection for human-robot fluency. *IEEE Transactions on Robotics and Automation* 23(5):952–961.

Hoffman, G., and Breazeal, C. 2008a. Anticipatory perceptual simulation for human-robot joint practice: Theory and application study. In *Proceedings of the 23rd AAAI Confererence for Artificial Intelligence (AAAI'08)*.

Hoffman, G., and Breazeal, C. 2008b. Achieving fluency through perceptual-symbol practice in human-robot collaboration. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction (HRI'08)*. ACM Press.

Hoffman, G.; Kubat, R.; and Breazeal, C. 2008. A hybrid control system for puppeterring a live robotic stage actor. In *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2008)*.

Jones, H., and Rock, S. 2002. Dialogue-based human-robot interaction for space construction teams. In *IEEE Aerospace Conference Proceedings*, volume 7, 3645–3653.

Khatib, O.; Brock, O.; Chang, K.; Ruspini, D.; Sentis, L.; and Viji, S. 2004. Human-centered robotics and interactive haptic simulation. *International Journal of Robotics Research* 23(2):167–178.

Kimura, H.; Horiuchi, T.; and Ikeuchi, K. 1999. Task-model based human robot cooperation using vision. In *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS'99)*, 701–706.

Komatsu, T., and Miyake, Y. 2004. Temporal development of dual timing mechanism in synchronization tapping task. In *Proceedings of the 13th IEEE International Workshop on Robot and Human Communication (RO-MAN 2004)*.

Levesque, H. J.; Cohen, P. R.; and Nunes, J. H. T. 1990. On acting together. In *Proceedings of AAAI-90*, 94–99.

Meisner, S., and Longwell, D. 1987. *Sanford Meisner on Acting*. Vintage, 1st edition.

Sebanz, N.; Bekkering, H.; and Knoblich, G. 2006. Joint action: bodies and minds moving together. *Trends in Cognitive Sciences* 10(2):70–76.

Simmons, K., and Barsalou, L. W. 2003. The similarity-in-topography principle: Reconciling theories of conceptual deficits. *Cognitive Neuropsychology* 20:451–486.

Weinberg, G., and Driscoll, S. 2006. Toward robotic musicianship. *Computer Music Journal* 30(4):28–45.

Wilson, M., and Knoblich, G. 2005. The case for motor involvement in perceiving conspecifics. *Psychological Bulletin* 131:460–473.