What Can Actors Teach Robots About Interaction?

David V. Lu* **Annamaria Pileggi†** **Chris Wilson*** **William D. Smart***

*Department of Computer Science and Engineering
†Performing Arts Department
Washington University in St. Louis
One Brookings Drive
St. Louis, MO 63130
United States
{wds,apileggi,davidlu}@wustl.edu

Abstract

In social interactions, the appropriate timing of physical actions is of vital importance. Otherwise appropriate actions performed at the wrong time can completely alter their perceived meaning. In this paper, we propose an approach to developing a physical vocabulary for human-robot interaction that uses trained actors as models of appropriate timing and gestures. We describe out initial work with robots in the performing arts, and discuss how this led us to our current approach, and why we believe that it will succeed in improving human-robot interaction.

Introduction

In social interactions, the appropriate timing of physical actions is of vital importance. Otherwise appropriate actions performed at the wrong time can completely alter their perceived meaning. Nodding in response to a speaker can signal understanding or agreement when done at the right time, just after the utterance. However, a slight pause might signal inattentiveness or hesitation, depending on the context of the interaction, and the content of the utterance.

The problem of timing is an acute one, since most humans are extremely sensitive to it. At the same time, it is hard to quantify precisely, because of its dependence on the current context (which is, itself, often hard to pin down).

However, most humans are able to generate actions with appropriate timing in a social interaction. It seems reasonable, then, to use humans as a model for our social robots. We can observe humans as they interact, note the gestures that they use with each other, time them, infer their cues, and replicate things on our robots.

There is a problem with this approach: many of the social gestures that humans use are subtle, such as a flick of the eyes, and current robot hardware is not capable of reproducing them accurately. Most pan/tilt units, for example, are not capable of moving fast enough to emulate the saccadic motion of human eyes. Worse still, the robot may completely lack the corresponding body parts.[1] For a human, a nod of the head and scanning down-and-up with the eyes can mean very different things. For a robot with only a traditional pan/tilt unit, these motions would have to map to the same bob of the cameras.

Our proposed solution to this problem is to use trained actors as models for our robots. Actors are trained to express themselves physically, and to make this expression clear and unambiguous to an audience. They are used to working under constraints, both of situation and of physical movement and, importantly for this work, are adept at repeating performances over and over, with minimal variation.

We have begun a collaboration that investigates the use of actors as models for action timing in human-robot interactions. In this paper, we describe some of our initial work in bringing actors and robots together, and how we plan to build on this to improve social human-robot interaction. However, before we describe our initial work, it is useful to provide some background on how actors are trained to think about movement and timing.

Related Work

There is strong evidence (Reeves & Nass 1996) that people treat non-animate objects as social actors, following social norms when dealing with, for example, computers. We claim that, by extension, this will be true for robots. In fact, we believe that it will be an even more powerful phenomenon with real robots, simply because of their physical embodiment. Thus, we need to give our robots appropriate social cues if they are to interact naturally with humans, even if the task of the robot is not inherently a social one. This view is borne out by Fincannon *et al.* (2004), who describe the social interactions that occur between search and rescue personnel and their robots.

Much of the work in expressive actions has been done in the realm of conversational agents. In these systems, an animated on-screen character holds a conversation with a human. Often the purpose of this conversation is to impart specific information, such as directions, and much of the work has focused on gestures by the character that help convey this information. This work is still relevant to our approach, however, since as Kidd & Breazeal (2004) point out, robots are similar to animations, but far more engaging because of their physical presence.

Cassell (2007) shows that body pose, hand gestures, and

Copyright © 2010, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

[1] Our research is primarily concerned with non-anthropomorphic robots. For human-like robots there will, by definition, be an appropriate mapping, although the Uncanny Valley (Mori 1970) may still cause problems.

eye-gaze direction signal changes in a conversation, initiating or terminating interactions, or facilitating turn-taking. Again, these observations lend support for the inclusion of such secondary actions in a mobile robot's repertoire. Further, it seems that the addition of these body-language cues reinforce the use of human conversational protocols (greetings, turn-taking, etc), and make them more effective (Cassell *et al.* 2001). This view is strengthened by Sidner *et al.* (2006) who found that users interacting with a robot penguin were more engaged, and used more head gestures, when the robot itself nodded appropriately. Further evidence is provided by Bruce, Nourbakhsh, & Simmons (2002), who note that robot head movements are much more likely to engage passers-by the simple vocalizations. However, they also report that breaking social norms in this situation is "unpleasant and unnerving" for the humans involved once these physical actions are in place.

Sidner & Lee (2007) provide more evidence that attentional gestures are vital to managing the flow in conversation. They note that such gestures fall into two categories: those that convey semantic information, and those that scaffold the conversation itself. In the latter category (which we are more interested in for our work) are head and eye movements, body stance and position, and some hand gestures. These so-called *engagement behaviors* allow the participants in a conversation to more effectively take turns, and to follow accepted social norms. There is also evidence to suggest that these gestures play a larger role than perceived emotional response for conversational agents (Cassell & Thórisson 1999).

Training the Actor in Physical Action

Everything we, as human beings, say and do has a purpose; some underlying need or goal that guides our behavior. Acting is a craft in which one explores communication in varying scenarios, involving people with conflicting needs and goals. Actors are trained to become more aware of their expressive potential, both verbally and physically. This heightened awareness is a necessary first step in effectively "acting" on another person to obtain some goal or objective. Here, we consider two rough categories of acting: pure movement and contextual acting.

Pure Movement

One method of physical training commonly employed is Viewpoints (Bogart & Landau 2005). Viewpoints is "a philosophy translated into a technique for (1) training performers; (2) building ensemble; and (3) creating movement for the stage" (Bogart & Landau 2005, p 7). This technique is comprised of interactive movement exercises that explore space and time. There are nine viewpoints, including five spatial elements: architecture, shape, gesture, floor pattern, and spatial relationship; and four temporal elements: repetition, tempo, duration and kinesthetic response. These viewpoints are a pedagogical tool to help the student actor think about their own movements, and how to respond to other actors' movements and the world around them. By including or omitting certain viewpoints, the actors can focus their attention on certain aspects of their movement and response, and more easily hone their technique. In an "Open Viewpoints" exercise (an advanced form of this training), the actor, through improvisation, uses any combination of these elements to influence and guide his interaction with other actors.

In the early phase of training, an emphasis is often placed on "pure" movement. This is movement without any intentional emotional context or agenda. It is movement for its own sake; a physical language of exploration. This agendaless exploration accomplishes several important goals. Most importantly, it fosters a wider range of physical expression than would be possible if the actor were performing in some pre-specified context. It also develops increased impulsiveness and openness, training the actor to react naturally to her environment and those around her. Finally, it enhances perceptual awareness, allowing the actor to take in all of the performers and environment around him, and to actively use this in his performance.

Students are introduced to this principle of "pure" movement through various exercises, one of which is The Flow. The Flow consists of five basic instructions. The actor can (1) start or stop; (2) change direction; (3) change tempo; (4) explore the space in between; and (5) follow. With this limited vocabulary, the actor is given a solid structure, enabling him to simplify his interactions to receiving a stimulus, having an impulse, and giving a response. In doing this, the actor is practicing the most basic approach to physical interaction. In time, this exercise becomes a useful warm-up, much like running, dribbling, and shooting exercises at the beginning of a basketball practice. It helps to center and focus the actor, preparing him for the more complex Open Viewpoints exercise which employs all nine of the spatial and temporal elements in improvisation.

Contextualized Movement

However, what most people consider acting does not fall into the "pure movement" category. Instead, as actors progress in their training, scenes are given the contextualizing elements of objective, relationship and place. The objective forces the actors to need something specific and to have a goal. The relationship requires them to be engaged in a specific relationship (such as siblings, friends, lovers, etc.), and the place requires them to place their actions in the context of the environment around them. These elements give the actor a new form of structure to guide physical interaction: movement becomes more than pure stimulus/impulse/response in time and open space. The three elements of context give an agenda to the movement. These "given circumstances" connect the actors and their actions in a more specific way. These factors make the movement active because it is now performed "to" or "for" the other person.

The idea of context is related best given an example of a more advanced exercise that introduces some initial contextual elements, while maintaining the simple array of movements introduced in the Viewpoints exercises. In this exercise, two actors perform a piece titled "Forgiveness", in which Actor A has recently betrayed Actor B. Actor A's objective is to seek forgiveness, while Actor B wants an apology. They are instructed to decide upon a specific relationship and locale for the piece. These elements must be conveyed to the audience without verbal text, using only physical action from the following vocabulary:

- move towards
- move away from
- sit/stand
- look at
- look away from
- embrace
- touch

In these exercises two things dictate an actor's movements: their objective and the movement of their partner. For instance, if an actor is seeking forgiveness, he might move towards his partner in an effort to appease his partner. As a response to this movement toward, the partner may look away from him in an effort to reject his attempt at appeasement. His response then may be to move away from his partner and sit down in order to give the partner some time and space. This physical "talking and listening" will comprise the "score" of the movement piece. Each actor must try to achieve his objective. The piece can end in one of two ways: (1) Actor A is forgiven, (2) Actor A is not forgiven. At this point one or both will exit the space.

The "Forgiveness" piece is based on a structure to introduce students to the "Method of Physical Action," developed by Polish director Jerzy Grotowski. According to Grotowski, contextualized physical action must "...have a why, a for whom, or an against whom." (Richards 1995, p 74). This is the actor's objective. The series of physical actions performed by the actors, i.e. the score, gives the actor a specific structure, a line of physical actions which comprise his journey through the piece. Each time the piece is performed, the actor repeats his score. However the need (objective) behind the score must be newly felt each time. The objective which necessitates the encounter between the two actors and the journey through the score must be experienced in the moment each time the piece is performed. This creates a seeming paradox for the actor:

> ...I was unaware that I was witnessing the two aspects so important to the creative process in theatre, the two poles that give a performance its balance and fullness: form on one side, and stream of life on the other, the two banks of the river that permit the river to flow smoothly. Without these banks there will be only a flood, a swamp. This is the paradox of the acting craft: only from the fight between these two opposing forces can the balance of scenic life appear. (Richards 1995, p 21)

Both structure and stream of life, i.e. the experience, must be present in performance. Without structure, there is nothing to contain the life, and without life, the structure becomes empty.

> Theatre is an encounter. The actor's score consists of the elements of human contact: "give and take." Take other people, confront them with oneself, one's own experiences and thoughts, and give a reply. In these somewhat intimate human encounters there is always this element of "give and take." The process is repeated, but always *hic et nunc*: that is to say it is never quite the same. (Grotowski 2002, p 212)

Figure 1: A scene from the performance of "Dr. Oddlust".

In terms more familiar to human-robot interaction researchers, the structure corresponds to the actions executed during the interaction and the "stream of life" is the timing of these actions. Even the correct actions are useless if their timing is wrong. It is this timing that breathes life into a performance, and makes the interactions in it seem natural.

Initial Work

In this section, we describe our experiences with putting robots into traditional theatre performances, using them in movement classes, and our first attempts at combining performance with quantitative user studies. In all of the studies, we used either an iRobot B21r mobile robot (pictured in figure 2), or a Videre Design Erratic ERA..

Robot as Actor

Our first use of a robot in a theatrical setting was as an actor in a student-run play festival at Washington University. The robot appeared in a short performance along with four human actors, and was seen by an audience of over one hundred people. The plot of the play, titled "Dr. Oddlust", involves a woman introducing her new boyfriend to two of her friends, with the twist that the boyfriend is actually a robot. The women debate whether it is proper for the relationship to exist, with one of the friends actually making a pass at the robot. At the end, it was revealed that the robot was not in fact the woman's boyfriend, but was stolen from a robotics lab. The play ends with the researcher who has come to retrieve the robot alone on stage with the robot, where her own secret affair with it is revealed.

The main goal was simply to see how an audience would react to seeing an actual robot (as opposed to one created specifically for the performance) in a theatrical setting. Based on the audience reaction, we were happy to discover that the audience accepted the robot in the role it was given, in a manner appropriate to the context of the play. The emotional responses of the audience indicated that the relationships between the characters were also accepted. Furthermore, despite its limited range of physical movement, the robot was, at times, able to communicate to the audience its internal state. We hypothesize that this is partially due to the very affective nature of the pan/tilt unit, which we found ef-

Figure 2: The robot taking part in a Viewpoints exercise in the “Fundamentals of Movement” class

fective at conveying attentiveness, sadness, as well as many other behaviors.

This collaboration resulted also gave us a useful experience working with actors in their domain. We realized early on that, for a good performance, the robot would need to be tele-operated from off-stage. Simply programming in a sequence of actions for the robot to progress through would have betrayed the essential “give and take” of theatre. Further, it would force the other actors into a very strict and unnatural timing of their own actions. Hence, The entire context of the robot’s relationships with each of the characters relies on how and when they react to each other. Hence, lacking much more advanced sensor models, tele-operation was essential to ensure the proper timing of the interactions.

Robot as Movement Partner

We brought our robots to two movement classes in the Performing Arts Department at Washington University in St. Louis taught by Pileggi. In both classes, students were briefly introduced to the robot as “an expert in Viewpoints”, and were told to interact with it in the same way as they interacted with their fellow actors. No technical details of the robot were given. None in the class admitted to prior experience with robots.

In the first session, approximately 20 students in the “Fundamentals of Movement” class participated in a Flow exercise (figure 2). The B21r robot was tele-operated from a control booth overlooking the room by Wilson, an experienced actor well-versed in the exercise. In the second session, an Erratic robot interacted autonomously with a small class of four experienced actors in a Flow exercise. A number of simple reactive behaviors were implemented on the robot, allowing it to interact with the actors and its environment in accordance with the rules of the exercise.

Part of the appeal of inserting the robot into the Viewpoints exercises is pedagogical. While it is initially difficult for human actors to remove context from their movements, it is much easier for them to assume the robot has no intention. When interacting autonomously, the robots movement’s are truly reactive and random. This gives the actors something close to an ideal model of how to perform the exercises. Furthermore, because its actions were so inherently different than what the actors were used to, they were forced to react to the movements in different ways themselves, ensuring that they were truly reacting “in the moment.”

However, despite the “otherness” of their new acting partner, many of the students began to trust the robot in the scope of the exercise. The students often referred to the robot in the same way they would refer to another human actor. Many called the robot a “he” and said he was “very present.” The actors all answered affirmatively on the question of whether they trusted the robot or not. Their willingness to treat the robot as an equal provides a steady basis on which we can build further interactions between robot and human actors. Furthermore, these exercises gave us an opportunity to observe the robot interacting with human actors in a highly controlled setting, allowing us to refine the timing of the robot’s interactions.

Performance and User Study

The experiments described in the previous sections, while informative, lacked a clear comparison between the robot and human actors. “Dr. Oddlust” would not have made sense with a human in the boyfriend role, and Viewpoints is highly variable in each running of the exercise. In this section, we describe a study directly comparing the robot to a human performing the same scene.

In order to integrate robots into theatrical settings and integrate acting techniques into robot behaviors, we need to determine how level the playing field is. In other words, if robots and actors are wholly incompatible and treated completely differently, then our hypothesis must be rethought. Hence, the perception of the robot in the eyes of an audience is key. If we want to explore the tools that actors use to create interactions, we must understand how those interactions are perceived differently (if at all) when a human or robot performs them. We tested this by producing three movement pieces, including the “Forgiveness” piece discussed earlier, in two different scenarios. In the first scenario, humans performed both roles, and in the second, the robot performed one of the roles with a human parter (see figure 3). All of the movement was restricted to movements that could be performed by the robot and a human in order to ensure that the performances were as similar as possible. The actors rehearsed the pieces to the point where they could reliably reproduce the movements and timings exactly. These movements and their timing were recorded, and replicated exactly on the robot. The robot was controlled autonomously for these pieces, running a pre-timed set of actions, to avoid any potential problems with trying to recognize cues from the actors using the robot’s sensors. Of course, this made it vital for the actors to stick to the rehearsed timing in order for the scene to work.

The pieces were performed before 33 people, over three shows. Each audience saw all three pieces with both the robot and a human actor in each piece in a random order, for a total of 6 performances. The audiences were told that

Figure 3: The same scene from two performances, one with two human actors (left) and one with one robot and one human actor (right).

they were going to see a collection of performance pieces. After each piece, the users were asked a series of questions about what they had seen. In the questions, the human actor in both plays was referred to as the "blue actor" and the part played by both the human and robot was referred to as the "red actor".

Although we did not have enough study participants to draw statistically significant conclusions, we did note some interesting trends in the collected data. When asked to ascribe a role to the red actor or a context to the piece, there was little difference between the pieces with the robot and human actor. This suggests that the audience reacted similarly to both the human and the robot when performing the same actions in the same context.

One notable difference was that the audience reported being more entertained by the pieces with the robot than those with the two human actors. Although we attribute this to the novelty of having a robot in a performance piece, we believe that it bodes well for further performances involving robots, since the purpose of theatre is to entertain its audience.

Each participant was asked to select words that they believed applied to the actors, from a short list. The list included words such as "happy", "sad", "angry", and others associated with emotion and affective state. In two of the three pieces, the human actor was seen as considerably more aggressive, confrontational, and angry than his robot counterpart. While the cause of this difference is not clear, it is interesting to note that the introduction of a robot affected the perceived tone of the interaction in the scene, even though all of the movements and timing were the same.

Despite the limited number of participants, we found no evidence that the audience treated the robot and the human actor differently, in the majority of cases. In fact, the audience response was surprisingly similar, both as they watched the performances, and in their responses to our questions. The audience was able to successfully attribute the same context and intent to the scenes, regardless of whether the robot or the human was in them. We see this as an encouraging result for our future work.

Building a Physical Vocabulary for Human-Robot Interaction

Our experiences with robots and the performing arts have led us to an approach for designing a physical vocabulary for human-robot interaction. By "physical vocabulary", we mean a set of actions, their timing, and the cues that trigger them, that allow a robot to communicate social information to a human interaction partner. For example, the set of actions that allows the robot to communicate that it is understanding what the human is saying to it, without interrupting her, would be part of this physical vocabulary. In this example, the actions are likely to include head nods, eye and hand movements, and short vocalizations. The timing and cues will be determined by the interaction, and might include performing a confirmation action at the end of each phrase that was understood.

Rather than saying "this context requires these actions", then implementing and evaluating them on a robot, we prefer the following approach. Give actors a scene, and allow them to act it out. Videotape this, and let people watch it, and assign a context and a meaning to the actions. Then, use the actions that get a consistent response as a physical vocabulary for the robot. Avoid actions that are ambiguous. Implement the vocabulary, and repeat the evaluations, making sure the same actions evoke the same responses. If they do, then we can use them. If not, we discard them.

More specifically, the steps in our approach are:

1. Write a scene that models a common human-human interaction. For example:

    One person (A) is explaining a long set of directions to another (B). B is interested in the directions, and seems to understand them as they are given to him. The scene ends when A is sure that B has understood the directions completely.

The scene should capture one or more aspects of human-robot interaction that we wish to implement. In this example, it is B's understanding of the directions, and the timing of the motions that signify this understanding, that we are interested in.

2. Have actors rehearse this scene, without any fine-grain direction, such as how to move, what expressions to wear, which gestures to make, etc. Their actions, however, will be constrained by the motions that our robot is capable of. Once the actors are happy with the scene, they perform it, and are recorded (on video and potentially by motion-capture).

3. Have an audience of "normal humans" watch the performance, either live or on video, and ascribe intent and context to the piece, and the specific actions that are being done. Essentially, this amounts to coding a video of the performance.

4. Pick the actions and timings that get a consistent audience response (in this context), and implement them on the robot. Avoid actions and timings that get an inconsistent response. Our goal is not to replicate the actor's performance; it is to replicate the *useful* aspects of the performance. For us, the useful aspects are those that allow an observer to correctly and consistently infer the context of the interaction, and the intent of the robot.

5. Rerun the performance, with the robot standing in for one of the actors, to verify that the reactions are the same, in the same context. Since we are not pre-supposing that our robot can effectively sense the interaction cues of its human partner, the actor's ability to perform with a given timing is vital here. It allows us to hard-code the robot's reactions for the purposes of evaluation. Although this is not sufficient for real situations, it does allow us to make progress on creating a physical vocabulary for the robot, without having to worry about the (extremely hard) sensor processing problem.

6. Retain the actions that evoke the same responses. Discard the rest. Identify each of the actions with the interaction context, and the set of cues and timings used to trigger them in the performance.

This, of course, ignores the problem of determining the context of the interaction that the robot finds itself in, and of recognizing and responding to the cues of the human partner. Although our approach will build a physical vocabulary to use for HRI, knowing when to use the elements of this vocabulary is left for future work.

Acknowledgments

This work was partially supported by the NSF under award IIS-0917199. "Dr. Oddlust" was written by Nick Loyal, directed by Andy Bird, and performed by Julia Martin, Carli Miller, Sarah Auerbach, and Ceng Chen. The user study pieces were written by Anna Pileggi, Justin Rincker and Louise Edwards, directed by Pileggi, and performed by Justin Rincker and Elizabeth Birkenmeier.

References

Bogart, A., and Landau, T. 2005. *The Viewpoints Book: A Practical Guide to Viewpoints and Composition*. Theatre Communications Group.

Bruce, A.; Nourbakhsh, I.; and Simmons, R. 2002. The role of expressiveness and attention in human-robot interaction. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2002)*, volume 4, 4138–4142.

Cassell, J., and Thórisson, K. R. 1999. The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence* 13(4 & 5):519–538.

Cassell, J.; Bickmore, T.; Campbell, L.; Vilhjálmsson, H.; and Yan, H. 2001. More than just a pretty face: Converstational protocols and the affordances of embodiment. *Knowledge-Based Systems* 14:55–64.

Cassell, J. 2007. Body language: Lessons from the near-human. In Riskin, J., ed., *Genesis Redux: Essays in the History and Philosophy of Artificial Intelligence*. Chicago, IL: University of Chicago Press. 346–374.

Fincannon, T.; Barnes, L. E.; Murphy, R. R.; and Riddle, D. L. 2004. Evidence of the need for social intelligence in rescue robots. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004)*, 1089–1095.

Grotowski, J. 2002. *Towards a Poor Theatre*. Routledge.

Kidd, C. D., and Breazeal, C. 2004. Effect of a robot on user perceptions. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 4, 3559–3465.

Mori, M. 1970. The Uncanny Valley. *Energy* 7(4):33–35. Translated by K. F. MacDorman and T. Minato. Original title: Bukimi no tani.

Reeves, B., and Nass, C. 1996. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York, NY: Cambridge University Press.

Richards, T. 1995. *At Work with Grotowski on Physical Actions*. Routledge.

Sidner, C., and Lee, C. 2007. Attentional gestures in dialogues between people and robots. In Nishida, T., ed., *Engineering Approaches to Conversational Informatics*. Wiley and Sons.

Sidner, C. L.; Lee, C.; Morency, L.-P.; and Forlines, C. 2006. The effect of head-nod recognition in human-robot conversation. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, 290–296.