Performer Model: Towards a Framework for Interactive Performance Based on Perceived Intention

Greg Corness
Simon Fraser University
1345 King Edward Highway
Surrey B.C. Canada
011 (604) 505-1108
gcorness@sfu.ca

ABSTRACT
Through the developing of tools for analyzing the performers sonic and movement-based gestures, research into the system-performer interaction has focused on the computer's ability to respond to the performer. Where as such work shows interest within the community in developing an interaction paradigm modeled on the player, by focusing on the perception and reasoning of the system, this research assumes that the performer's manner of interaction is in agreement with this computational model. My study presents an alternative model of interaction designed for improvisatory performance centered on the perception of the performer as understood by theories taken from performance practices and cognitive science.

Keywords
Interactive performance, Perception, HCI

1. INTRODUCTION
For the past two-decade, composers have been designing interactive music systems that are often viewed as new musical instruments, or as an emulation of a player or conductor [10][3]. As processor speeds increase, the systems being designed not only produce more complex sounds, but generating responses and analyze performer's gestures with increasing sophistication. In conjunction with these developments, increasing amounts of intelligence and autonomy are being built into systems for use in a variety of performance situations including improvisation. But as the autonomy of these systems increases it may be necessary to reconsider the models used for designing the interaction. Research into the performer-system interaction has focused largely on the computer's ability to respond. As composers explore giving agency to the computer, the performer is being required to be responsive. This study addresses a number of issues that lead towards constructing a framework for developing a performer-based model for improvisatory interaction.

2. TRADITIONAL MODELS
In the early years of interactive music, Robert Rowe proposed a distinction between an instrument paradigm and a player paradigm as one axis along which we could place different interactive systems [7]. Rowe suggests, " Instrument paradigm systems are often more concerned with timbral generation, while the player paradigm requires the use of some meta-compositional generation method to produce musical output" [10]. His taxonomy, though mainly focused on process used to generate the system's response, implies a consideration of a fundamental difference in the manner of interaction. The instrument paradigm suggests devices used for direct control of synthesis and low-level parameters (pitch, volume, on/off), while the player paradigm generally involves sensors that allow for the mapping of larger performative gestures to global parameters.

Improvisational music systems often implement elaborate sensors and algorithms for analyzing the physical and sonic gestures of the performer [7] [14] [11]. The assumption underlying this approach is that much of the communication between performers, and in particular musicians, is through the context and syntax of their sonic response. This argument is not wholly untrue and has produced some very accomplished systems. However, musicians tend to play within what might be termed social contact with each other. With this term I refer to communications modes, such as eyesight, that are separate from the act of playing, but I also intend to bring attention to the social aspects of music that give it common ground with other performance disciplines.

Other interactive music projects have expanded the mode of communication to explore other cues such as visual movement cues [14], acoustic variation [7][14] and multisensory multimedia [3] often in the context of interdisciplinary performance. Exploration in Multimedia and interdisciplinary interaction has found that the system-performer communication cannot rely on the syntax of a particular performance domain but rather be expanded to general expressive gestures.

This research into performer-system communication shows an interest in developing a player paradigm for interaction. However, the research has focused on the perception of the system thus ignoring aspects of human communication. This study presents an alternative model of interaction appropriate for improvisatory performance by examining theories taken from performance practices and cognitive science to focus on the performer's ability to perceive intention.

3. BACKGROUND THEORY
Communication in performance is an inter-subjective phenomenon where understanding is *agreed* upon by the agents involved in the moment. As Lockford and Pelias explain:

> "Even when faced with the challenge to perform in an unscripted moment, performers understand that they are engaged in an ongoing communicative exchange. This exchange is a process best conceived, not as an act of information transmission or shared understanding, but as communication scholar H. L. Goodall, Jr. would have it, as an act of 'boundary negotiation'." [8, p433]

Here "boundary negotiation" refers to the process of the self of the performer being incremental build with in the context of the performance. In a theatrical sense, this is the build up of character as new information is reveled in the scene. In a musical sense, the negotiation is between soloist and accompanist over harmonic extensions and rhythms that occur during a particular solo. Such a negotiation implies that the agent must be able to respond to new information while simultaneously presenting information to contribute to the self of other agents. Negotiation in these terms is a coordination of the interaction between agents [8]. It becomes imperative that all agents are able to negotiate the coordination of their intention and therefore able to track the intention of the others.

The importance of the agent's ability to track intention can be clearly seen when considering the notion of trust. Since the agents constitute them selves and each other through the negotiation of boundries [8], this inter-subjective communication requires a sense of trust. For a performer to be open to constituting their performance identity anew in negotiation with others on the stage, they must trust the environment. Furthermore, a sense of support is established when their actions both affect and support other agents. Again, this support comes from trust in the inter-subjective understanding of the moment. This understanding keeps the ensemble synchronized, but requires that all the performer-agents are able to track the intention of the others. Therefore, it becomes imperative that all agents be able to project their own intentions.

3.2 Agencies and State Knowledge

Bogart and Landau coach students of improvisation to "*trust* in letting something occur onstage, rather then making it occur" [1]. Applicable to both sonic and physical gestures, their statement does not mean that nothing should be started but rather to avoid forcing a start. We might call this an additive approach where additive suggests that the agency is added to the state of the system whether it is in steady state or a dynamic state. The implications of this view can be seen when considering the response of the performer rather then the system. To trust in the something that will happen is to coordinate the actions, adding to action of the system. This cannot be done in response. The improviser must move beyond the cognitive and trust in the intuitive [8].

3.3 Intuition and Intention

Research in the field of neuroscience has recently suggested links between *intuition* and *intention*. Neurons found in pre-motor areas of the brain have been shown to fire not only when producing a sound or action, but when the subject hears the sound or observes others doing the action as well [5] [9] [6]. The firing of these neurons allows the subject to predict the outcome of their own actions as well as the actions of others. "This implicit, automatic, and unconscious process of motor simulation enables the observer to use his/her own resources to penetrate the world of the other without the need of theorizing about it" [4]. What is crucial to this phenomenon is that the action observed must be goal oriented, that is it must have intention [5][9][6].

However, there is some question as to the usefulness of mirror neurons in human-computer interaction. The findings to date concerning a person's ability to perceive intention in others suggest that the ability diminishes in correspondence to the physical similarity with the other. This means that a human subject perceives the intention of other humans, but less so apes, only slightly with other animals and not at all with machines [4] [5]. The prevalent reason given for this distinction is a perceived similarity of motion [5]. It is then unclear whether a system's response actions would affect the per-cognitive process of a subject if accurately modeled on human action.

Still, the presence of the pre-cognitive function implies that the human cognitive system as a whole works in connection with this mechanism, and that even at a cognitive level, interaction is governed by the prediction of events as much or more then reaction to events, an interpretation supported by the presented theories on improvisational performance. These findings suggest that as social being we have developed the ability to intuitively predict the actions and sounds of those around us.

The idea that human action and intention happens before the act has been shown in other experiments as well. Wegner in his book "The Illusion of Conscious Will" presents the work of Kornhuber and Deecke (1965) as well as Libet (1983). These researchers measured a rise in brain activity up to 800ms before an action took place. In the case of Libet's experiments, brain activity was recorded over 300ms before the subject was even aware they wanted to act [13].

These findings further indicate that humans do not live in a static present moment but rather in a moment becoming the next. Our social engagements are informed by an embodied empathy that allows minor predictions of those around us. We react not in the moment but in the moment next over half a second late.

4. PERFORMER MODEL

The theories presented give an understanding of the role of perception of intention in human interaction. Based on these theories, I suggest that a framework for interaction between autonomous agents should address:

1) The need to negotiate boundries and build trust with others.
2) The development of an inter-subjective understanding of the moment
3) The need to feel supported through one's agency and acceptance in the environment.

I propose that these criteria may be address by incorporating in to the system a mechanism to allow the performer to perceive the system's intention. Therefore, I have started a series of studies looking at the experience of the performer working in a system designed to project its intention.

5. SYSTEM DESIGN

The system used to conduct the study took two forms, visual and sonic. Both systems were constructed through an iterative design process using a first person methodology. In order to focus the study on methods for modeling an embodied projection of the system's intention, gestures in both systems were generated with simple random processes, avoiding any signifiers that may come from structure or syntax, and allowed the system to enact its own "intention" with no sense of the performer. The response paradigms chosen for both test systems were informed by human response and perception behaviors but were not meant to mimic them. Finally, the research was set up as studies into the experience of a subject being afforded the ability to move with the system. No expectation of creation or performance was imposed.

5.1 Visual System

The response gestures in the visual system were realized using an image of two concentric circles generated in MAX/Jitter. This image was projected onto the floor of the performance space using an I-CUE dmx controllable mirror. The behavior of the system was set so that the inside circle needed to move off center for the entire image to move in the space. Stopping required the circle to return to the center. The direction and amount that the circle moved off center corresponded to the direction and speed at which the image was about to move. The time required for the inner circle to reach its maximum point was set at 200ms, in line with the research presented by Wegner. The movement of the light object was constrained using a dynamic weighted random algorithm. The probability of the light moving in any direction was a function its position in the space.

5.2 Sonic System

The sonic version of the study was modeled on the common idea that breath can be used to synchronize a group. The system used a physical model of a flute constructed in the PeRColate synthesis library for MAX/MSP [12]. Each session explored different approaches to perceiving information embedded in different parts of the breath sound. The information was embedded by manipulating the parameters of the flute model to get different qualities breath sounds before and after the tone. The timings of these different breath qualities in each session were functions of the generated gesture's length, density and speed.

6. QUALITATIVE DATA

6.1 Visual System

I spent a number of sessions working in the system to feel the experience of being in the space with it. As might be expected, it was easy to anthropomorphize the light. I Perceived it's motion as a nervous exploring intention, even though I *knew* the movements were random. Still, it quickly became apparent that the system had no sense of my presence. This had been part of the design, however, it was interesting to note how easily I perceived the design as experience. Furthermore, this perception profoundly changed the quality of the interaction from the intended design model of *tag* to one of playing in ocean waves or taunting a blindfolded partner. My perception of the system's movement intention, stalking and lunging with no focus on me, inspired a sense of teasing. I noticed myself considering which way the system was "thinking of moving" and circling to the other side just out of "reach". The random process used for starting and stopping also produced occasional motions perceived as "fakes" in which the Light Actor moved it's "weight in one direction then immediately moved it back to a center position. This emergent behavior was of special interest. The perception that I could tell where it was "thinking" of moving encouraged me to get close but the impression that it could "change its mind" kept up my interest in the engagement.

6.1.1 Test with non-projecting system

Some time was spent comparing the system with and without the center circle active. With out the center active I noticed I was not inspired to get close to the light, and my willingness to engage with the system was shorter. Similarly, I noticed when the response behavior was tuned to give less fakes the movements became easier to predict, but the interaction became less engaging in the context of a tag paradigm.

6.1.2 moving with the light

During a second session I focus on moving with the light rather then avoiding the light. At first I changing only my behavior, the system's behavior pattern remained the same as before; however, I found this interaction very unsatisfying. Although I could tell where the light was going, I had very little time to coordinate my own movements. The interaction quickly became a dodging rather then a moving with.

The behavior settings of the system were then changed to generate movements that tended to be longer with less "fake" motions. These changes were modeled after mirroring exercises in which human partners try to mimic each other's motion with out a sense of leading. In these exercises, fluid, often slow predictable motions are emphasized. With the system's behavior modeling mirror exercises, I found the interaction with the light more of a *moving with* experience. However, the quality of my movement remained at a "proof of concept" level. The interaction did not inspire flow or exploration in my own movement.

6.1.3 shape

As final note, I noticed that the circle inside a circle design had more the top down look of a joystick then a human. I tried giving a more human shape by using ovals rather then circles, but found the oval shape less engaging then the circles. Though this can be explained by the fact that an oval implies a direction and the system was not programmed to take direction of the image into account, my experience suggests that the circle configuration, though endowed with behavioral characteristics, remained a spot of light. My perception of the object combined "lightness" with behavior and did not need to construct a new humanoid entity.

6.2 Sonic System

The audio-based system had a different initial impact. Where as the visual system had inspired an avoidance response and only after being re-modeled, produced a *moving with* response, my experience was that the breath model in the audio based system immediately inspired a *moving with* response. The randomness of the gestures had less of an affect, perhaps because there were no fake gestures produced by the sonic system. The breath sound in the first session was linked to the duration of the generated phrase and produced a feeling of lift into tone of the sound. This feeling of lift encouraged my motion with the onset of the sound even though I had no knowledge of when it would happen.

Through reflecting on my response I noticed two parts to the breath generated by the physical model: the inhale and the stream focusing. I was lifting on the inhale but moving on the focusing change of breath just before the flute tone. This discovery inspired a series of sessions exploring the breaking of the breath sound into three parts: inhale, focused –airstreem and breath trail-off. By considering that a breath into a beat is often used to signal a down beat and that more air is needed to play longer phrases I mapped inhale duration to tempo and inhale volume to phrase duration. This mapping frequently allowed me to anticipate the tempo of the phrase and move with it but only with in a small range of values. However, when inhale duration was a function of phrase length I found that I moved with out much thought with the sound. The mapping of duration to tempo affected in me a more rational approach to moving.

7. DISCUSSION

The literature and theories presented in this paper suggests that human interaction is not restricted to reacting to enacted events. Instead, as social being, our interactions include the understanding and prediction of events through the perception of the intention of others in the environment. From these theories, I have suggested a framework for interaction, modeled around the abilities and needs of a Performer pertaining to perception of intention. The crucial point is that all agents in the environment need to be able to perceive the intentions of the other agents. The framework that I am constructing has a crossover with the "Player" paradigm of interaction, first suggested by Rowe, in that agency is being given to the system. However, the proposed framework differs from Rowe's paradigm by focusing on interaction through the perception of interaction rather then through a process for responding.

In order to demonstrate the implications of this approach in the context of both sonic and physical interaction, I have discussed two example systems: a visual based system and a sonic based system. Both systems were designed around the claim that the performer needs to be able to perceive the intention of the system in anticipation of any action. The result of my studio work indicates that both visual and sonic systems provide the opportunity to embed information in the system's response media, projecting the system's general intention. Analyses of the results indicate further that the two systems share many of the same issues. The cognitive load imposed on the performer when trying to predict the action of the system was reviled as an issue when using analytical models to indicating intentions. These models were most prevalent in the sonic system, and yet, a similar effect was observed in the visual system. Both systems indicated an experiential difference between "natural" and analytical interactions; however, the parameters for separating these qualities have not been isolated. What was made clear by the studio work was that the manner in which the system expressed its intention did not need to be "true", modeled on a human gesture. However, there is some indication that a stronger reference to signifiers that are already part of the performer's body knowledge reduced their need to rationally analyze the intention of the system. Of prime importance was the observation that a feeling of *trust* and *sharing of space* was created in the system projecting its intention that was not present in the response only system. With more investigation it is hoped that a system may be developed, that enables the integration and alignment of both the performer and system's intentions for a more unified and balanced interaction.

9. REFERENCES

[1] Bogart, A., and Landau, T., The Viewpoints Book: A Practical Guide to Viewpoints and Composition. *New York: Theatre Communications Group,* (2005)

[2] Camurri, A., and Feffentino, P., Interactive Environments for Music and Multimedia. *Multimedia Systems 7: 32-47* (1999)

[3] Camurri, A., et al. The MEGA Project: Analysis and Synthesis of Multisensory Expressive Gesture in Performing Art Applications. *Journal of New Music Research, 34:1, 5-21*

[4] Gallese, V. The "Shared Manifold" Hypothesis: From Mirror Neurons to Empathy. *Journal of Consciousness Studies 8 5-7* (2001) 33-50

[5] Gallese, V., The Intentional Attunement Hypothesis: The Mirror Neuron System and Its Role in Interpersonal Relations. *Biomimetic Neural Learning* (2005) 19-30

[6] Iacoboni, M. et al. Grasping the Intention of Others with One's Own Mirror Neuron System. *PLoS Biology 3:3* (2005): 529-35

[7] Lewis, G. E., Interacting with Latter-Day Musical Automata. *Contemporary Music Review 18:3* (1999): 99-112

[8] Lockford, L. and Pelias, R., Bodily Poeticizing in Theatrical Improvisation: A Typology of Performative Knowledge. *Theatre Topics 14.2* (2004) 431-43

[9] Kohler, E., et al. Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons. *Science 297,* (2002) 846-8.

[10] Rowe, R. Incrementally Improving Interactive Music Systems, *Contemporary Music Review 13:2 p.47-62 (1996)*

[11] Thom, B., Artificial Intelligence and Real-Time Interactive Improvisation. *Proceedings of the Seventeenth Conference on Artificial Intelligence, Austin Texas,* August (2000)

[12] Trueman, D., and DuBois, R. L.*, PeRColate: A Collection of Synthesis, Signal Processing, and Video Objects for MAX/MSP/Nato. V 1.0b3*

[13] Wegner, D. M., The Illusion of the Conscious Will, *Cambridge MA. USA: MIT Press* (2002)

[14] Weinberg, G., and Driscoll, S., The Perceptual Robotic Percussionist- *New Developments in Form, Mechanics, Perception and Interaction Design 2nd ACM/IEEE International Conference on Human-Robot Interaction Washington DC, USA,* March 9-11 (2007)