Designing for Conversational Interaction

Andrew Johnston
Creativity & Cognition Studios
Faculty of Engineering and IT
University of Technology, Sydney
andrew.johnston@uts.edu.au

Linda Candy
Creativity & Cognition Studios
Faculty of Engineering and IT
University of Technology, Sydney
linda@lindacandy.com

Ernest Edmonds
Creativity & Cognition Studios
Faculty of Engineering and IT
University of Technology, Sydney
ernest@ernestedmonds.com

Abstract

In this paper we describe an interaction framework which classifies musicians' interactions with virtual musical instruments into three modes: instrumental, ornamental and conversational. We argue that conversational interactions are the most difficult to design for, but also the most interesting. To illustrate our approach to designing for conversational interactions we describe the performance work *Partial Reflections 3* for two clarinets and interactive software. This software uses simulated physical models to create a virtual sound sculpture which both responds to and produces sounds and visuals.

Keywords: Music, instruments, interaction.

1. Introduction

We are concerned with the development of interactive software for use in live performance which facilitates what we call 'conversational' interaction. We work with expert musicians who play acoustic instruments and are intrigued by the potential of interactive technologies to provide new perspectives on sound, performance and the nature of interaction. While the term is imperfect and a little clumsy, we call the various pieces of software we have developed 'virtual musical instruments' or, more simply, 'virtual instruments'.

In this paper we present the findings from a qualitative study of musicians' interactions with virtual instruments we have developed previously and describe how these influenced the artistic direction of subsequent creative work, somewhat unimaginatively entitled *Partial Reflections 3*

2. Physical Models as Dynamic Intermediate Mapping Layer

The virtual instruments described in this paper have the following characteristics:

- Acoustic sounds captured via microphone are the source of 'gestures' which act upon the virtual instruments.
- These musical gestures result in force being applied to a software-simulated physical model (or 'mass-spring' model) which responds by moving in physically plausible ways.
- The movements of the simulated physical model provide parameters for sound synthesis.
- A representation of the physical model is shown on-screen, visible to both performers and audience. From their point of view the physical model *is* the virtual instrument.

This approach draws heavily on that described by Momeni and Henry [1] and Choi [2]. Audio input from the user results in force being exerted on the physcal model and in response parts of the model move about, bump into each other, etc. Various measurements of the state of the model, such as speed of individual masses, forces being exerted, acceleration and so on, are then separately mapped to parameters for the audio and visual synthesis engines. The visual synthesis mapping layer maps the X, Y and Z coordinates of masses to the position of geometric shapes on screen and the audio synthesis mapping layer maps characteristics of the masses (speed, force, etc.) to various synthesis parameters (such the individual amplitudes of a set of oscillators for example).

It can be seen that with this approach we end up with three mapping layers. The first maps from user gestures to parameters which change the state of the physical model. The second and third map from measurements of the state of the physical model to audio and visual synthesis parameters respectively.

This approach provides a number of advantages. Firstly, because both audio and visual synthesis parameters have the same source (the physical model), the intimate linkage of sound and vision is greatly simplified. While they may be separated if desired (by treating the outputs from the physical model in dramatically different ways), the 'default' condition is likely to lead to clearly perceivable correspondences between sound and vision.

Secondly, the dynamic layer provides convenient ways to build instruments based on divergent (one-to-many) mappings [3, 4] . A mass-spring physical model which contains

a network of say 10 masses linked together with springs can be set in motion by moving only one of the masses. The movement of this single mass results in multiple movements in the overall structure as the force propagates through the network via the links. Because the model applies the laws of Newtonian physics each of these movements is predictable at a high level and is a direct result of the initial user action. These derived movements provide extra streams of data which may be mapped to audio/visual synthesis parameters. [1]

Third, if the visual display is a representation of the dynamic layer itself (eg. a display of the actual physical model), then the user is more able to understand the state of the system, leading to an improved ability to control the instrument. In addition, such a display can help an audience understand and engage with a live performance as they are more able to perceive what impact the actions of the instrumentalist have on the virtual instrument.

Finally, the movements of the physical model bring a sense of dynamism to the virtual instrument. As the physical model network reacts to energy supplied by the performer it will often oscillate, providing rhythms the player can respond to. By bringing a sense of unpredictability and a kind of simple agency to the interaction, while still retaining high-level controllability, a physical model mapping layer may help stimulate a more conversational style of musical interaction [7]. We will return to this point later.

3. Modes of Interaction

Before describing the virtual instrument developed for *Partial Reflections 3*, we will firstly describe the interaction framework which provided the foundations for its design. In order to examine musicians' experiences with virtual instruments of the kind we describe here, we conducted a series of user studies. It is important to stress that we consider these user studies to be much more than exercises in evaluating the software instruments. More significantly, they are also investigations into the experiences of the musicians who used them. While we are interested in learning about the strengths and weaknesses of the virtual instruments, we are equally interested in the impact they have on the way the musicians make music. The virtual instruments are used to provoke current practice and in this sense they are 'provotypes' or provocative prototypes [8].

We had seven highly experienced, professional musicians (including principal players from symphony orchestras and leading jazz musicians) use three virtual instruments which used a simulated physical model as an intermediate mapping layer. (These instruments are described in [9].) The musicians were given minimal instruction regarding the virtual instruments, such as how they responded to the pitch and volume of their acoustic sounds, [2] and then given freedom to experiment with them as they pleased. They were asked to verbally report and reflect on their experience as they did so. In addition, a semi-structured interview was conducted in which musicians were asked to comment on various characteristics of the virtual instruments and their impact on their playing. Each session was video recorded and these were later transcribed and analysed using grounded theory techniques [10, 11].

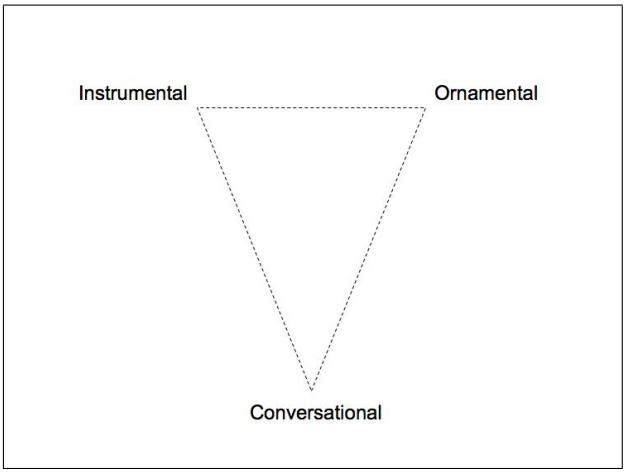

Figure 1. Three modes of interaction mark boundary points on a map of a musician's interactions with a virtual instrument.

The results of this study are reported elsewhere [12], but we summarise some of the key findings here in order to show how they influenced the design of *Partial Reflections 3*. A core finding was that the musicians' interactions with the virtual instruments could be grouped into three modes: instrumental, ornamental and conversational. These modes are not exclusive in the sense that one musician always interacted with the virtual instruments in one mode, or that each virtual instrument was only used in one mode. Some instruments did tend to encourage particular interaction modes but not exclusively. These modes of interaction could best be seen as boundary points on a map of an individual's interactions with a particular virtual instrument (figure 1). As such, a musician may for example begin in 'instrumental' mode, move to 'ornamental' mode for a time, and then eventually end up in a 'conversational' interaction.

Each of these modes of interaction will be briefly described in the following sections.

3.1. Instrumental

When approaching a virtual instrument instrumentally, musicians sought detailed control over all aspects of its oper-

[1] The Web, a physical controller designed by Michel Waisvisz and Bert Bongers [5, 6] also explores the interconnection of individual controller elements. The web is "an aluminium frame in an octagonal shape with a diameter of 1.20m., and consisting of six radials and two circles made with nylon wire" [6, p.63]. Tension in the strings was measured by custom designed sensors, providing a stream of data for sound synthesis.

[2] Some musicians preferred to use the virtual instruments without prior instruction, in which case this step was skipped.

ation. They wanted the response of the virtual instrument the be consistent and reliable so that they could guarantee that they could produce particular musical effects on demand. When interacting in this mode, musicians seemed to see the virtual instruments as extensions of their acoustic instruments. For these extensions to be effective, the link between acoustic and virtual instruments had to be clear and consistent.

3.2. Ornamental

When musicians used a virtual instrument as an 'ornament', they surrendered detailed control of the generated sound and visuals to the computer, allowing it to create audio-visual layers or effects that were added to their sound. A characteristic of ornamental mode is that the musicians did not actively seek to alter the behaviour or sound of the virtual instrument. Rather, they expected that it would do something that complemented or augmented their sound without requiring direction from them.

While it was not always the case, it was observed that the ornamental mode of interaction was sometimes a fall-back position when instrumental and conversational modes were unsuccessful. While some musicians were happy to sit back and allow the virtual instrument to provide a kind of background 'sonic wallpaper' that they could play counterpoint to, others found this frustrating, ending up in an ornamental mode of interaction only because their attempts at controlling or conversing with the virtual instrument failed.

3.3. Conversational

In the conversational mode of interaction, musicians engaged in a kind of musical conversation with the virtual instrument as if it were another musician. This mode is in a sense a state where the musician rapidly shifts between instrumental and ornamental modes, seizing the initiative for a time to steer the conversation in a particular direction, then relinquishing control and allowing the virtual instrument to talk back and alter the musical trajectory in its own way. Thus each of the three modes of interaction can be seen as points on a balance-of-power continuum (figure 2), with instrumental mode at one end (musician in control), ornamental mode at the other (virtual instrument in control) and conversational mode occupying a moving middle ground between the two.

To us, this implies that virtual instruments which seek to support conversational interaction need also to support instrumental and ornamental modes.

3.4. Discussion

The interaction framework we present here differs from other well known taxonomies of interactive music systems such as those proposed by Rowe [13] and Winkler [14] in two important ways. First, the modes of interaction were derived from a structured study of musicians. Rowe and Winkler's, in contrast, arose from their considerable experience designing and using new musical instruments. We certainly do not

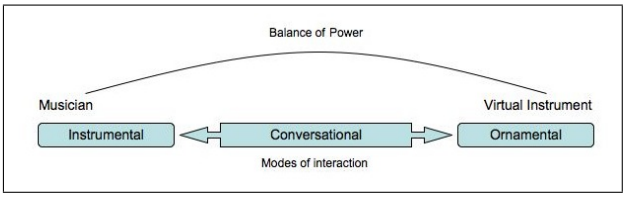

Figure 2. Virtual instruments which support conversational interaction facilitate a shifting balance of power between musician and virtual instrument.

suggest that our approach is superior, but we do point out that studies of the kind we have conducted can compliment personal experience reports and can be valuable in generating new perspectives. Second, our study focused on the experiences of the musicians who used the systems, as opposed to characteristics of the systems themselves. Studies of the kind we have conducted consider technical aspects of the virtual instruments in the context of the impact they have on the experiences of the musicians who use them. In this way they help to bridge the gap between system features and player experience.

4. Partial Reflections 3

In section 2 we described a technique for using simulated physical models as an intermediate mapping layer between live sound and computer generated sounds and visuals. In section 3, three modes of interaction which characterised musicians' interactions with virtual musical instruments which use this interaction style were briefly described. In this section, the design of a new virtual instrument, tentatively titled *Partial Reflections 3* (PR3) is described.

4.1. Context

As with all our instruments, PR3 was designed for use in live performance in collaboration with expert musicians, in this case the clarinetists Diana Springford and Jason Noble. The intention was to create a virtual instrument which would respond to the sounds of both players simultaneously but also independently: that is, the musicians would have separate channels through which they could act upon the virtual instrument, but they both interacted with the one instrument. The idea was that part of the musicians' musical conversation would be mediated by the virtual instrument, and that the virtual instrument itself would facilitate conversational interaction with the musicians. We were not interested in supporting purely instrumental or ornamental interactions.

Physically, the work was presented in a club-like music venue. The musicians flanked a screen which showed the visual output of the software. Their acoustic sounds were not amplified.

4.2. Technical Description

The simulated physical model at the core of *Partial Reflections 3* was comprised of 48 masses arranged in a large circle (figure 3). Each of the masses was linked to its neighbour

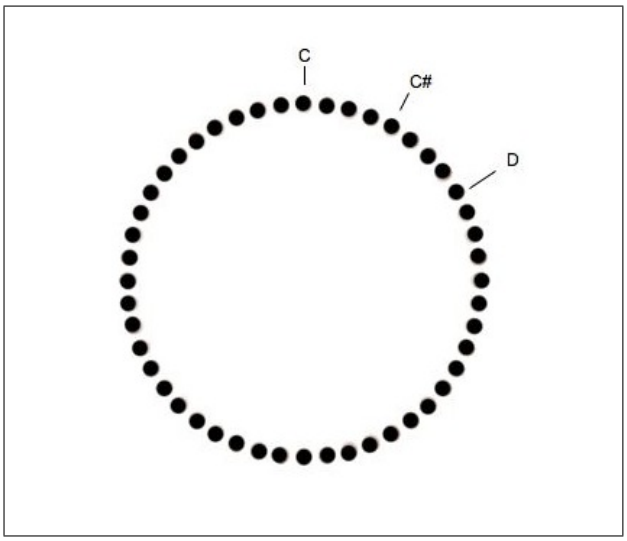

Figure 3. The physical model for PR3 was made up of 48 masses arranged in a circle.

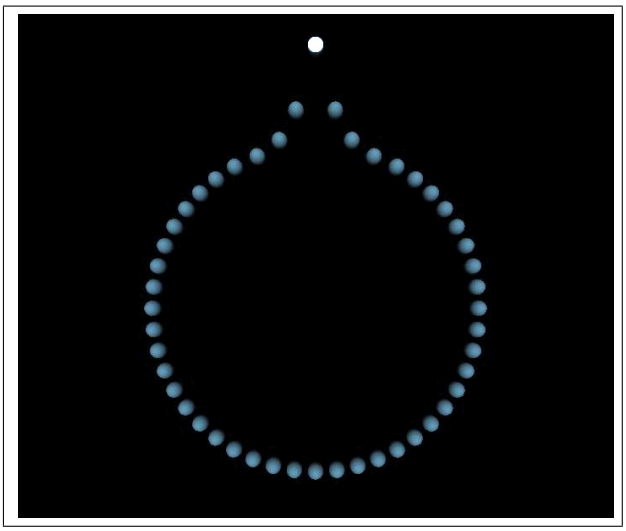

Figure 4. Screenshot showing effect on the physical model when a middle C is sounded on an acoustic instrument.

masses. In addition, in order that the masses remained in a circle, each mass was linked to an invisible mass which was fixed in position. [3] Finally, links were put in place which acted only when masses were effectively in contact with one another. The effect of this was to allow masses to bounce apart when they collided with one another.

The simulation itself was developed using Pure Data [15], GEM [16] and the Mass-Spring Damper (msd) object by Nicolas Montgermont. [4] Some helper objects written in Python were also used when the visual programming style of pure data was found unnecessarily clumsy.

In essence, the physical model acted as both visualisation of the musicians' acoustic sounds and as a controller for additive re-synthesis of those sounds. The computer-generated sounds could therefore be seen as a kind of echo of the live sounds mediated by the physical structure of the model.

The fiddle~ object [17] was used to analyse the audio streams coming from the two microphones. This was used to provide continuous data streams containing:

- Current volume.
- Estimated current pitch (and derived from this, pitch class).
- The three most prominent peaks in the harmonic spectrum.

The current volume was mapped to the amount of force exerted on the physical model and the current pitch class determined which of the 48 masses would be the target of that force. In order to map the octave onto 48 masses we simply divided each semitone by 4. That is, the mass at the top of the model was associated with the pitch class C, the mass immediately to its right with a C an eighth tone sharper than C, the next mass to the right with C a quarter tone sharper and so on around the circle. Thus, every fourth mass would be associated with a pitch-class from the standard 12 tone equal temperament scale (see figure 3). Forces always acted in an outward direction, pushing masses away from the centre of the circle.

An example should help to illustrate how this worked in practice. If a musician played a concert C on their acoustic instrument, the mass at the top of the physical model (ie. at the 12 o'clock position) would have force exerted on it. The amount of force would be proportional to the volume of the sounded note. In response the C mass would be pushed outwards from its resting position while the note was sounding (figure 4) [5] . Because each mass in the model is linked to its neighbour masses, the masses closest to the C mass are also dragged out of their resting positions.

Additive synthesis was used to generate sounds controlled by the movements of the physical model. In additive synthesis, complex sounds are produced by combining a number of simple waveforms - typically sine waves [18]. The pitch of the note played by the musician (ie. the frequency in Hertz) was mapped to the frequency of an oscillator associated with each mass. Because the model had 48 masses, there were 48 oscillators. If the musician played an A with a frequency of 440Hz (A above middle C) then the 'A' mass oscillator was set to oscillate at that frequency. If they subsequently played an A an octave lower (220Hz), then the 'A' mass oscillator was then set at 220Hz rather than 440Hz. The frequencies of the three strongest partials in the live sound were mapped

[3] If this was not the case then the floating (ie. non-fixed) masses would drift away from their starting positions in the circle as soon as forces were applied.

[4] http://nim.on.free.fr/index.php?id=software

[5] In order to aid transparency of operation, the mass which was currently having force exerted upon it was also made to glow.

similarly. If the 'A' played by the musician had strong partials at frequencies with pitch classes of E, G and C#, then the oscillators associated with those pitch classes were set to the frequencies of those partials.

Data from the physical model was used to control the output of the oscillators. The speed of each individual mass was mapped to the volume of its associated oscillator. The faster the mass moved, the louder the output from its oscillator.

4.3. Encouraging conversational interaction

As discussed in section 3, we believe that virtual instruments which support conversational interaction must fulfil the seemingly contradictory requirements of providing both detailed, instrumental control and responses which are complex and not entirely predictable. In order to facilitate instrumental interaction, the mapping between the acoustic sounds played by the musicians and the forces exerted on the physical model remained consistent during performance. That is, playing a middle C would always result in force being exerted on the C mass, for example. Likewise, the mappings between the movement of the physical model and the sounds produced by the additive synthesis engine were unchanged during performance. This helped ensure that the effect of performer actions on the virtual instrument could be predicted; if the musician played two perceptually identical notes on their acoustic instrument, the effect on the virtual instrument would be the same.

This is not to say that the *response* of the virtual instrument would necessarily be the same however. One of the consequences of using physical models as a mediating mechanism between performer gestures and virtual instrument response is that the response of the virtual instrument to a given musical input will change over time. That is, two perceptually identical notes played at different times during the performance may cause the virtual instrument to move in different ways (and therefore produce different sounds). This is because the state of the physical model changes over time. The physical model starts in a resting state and when a note is played it moves as a result of force being exerted upon one of the masses. If the same force is exerted on the same mass before the model has returned to its resting point, the response of the virtual instrument will be different to when it was at rest, because the model is in a different state.

The response should be predictable to musicians however, because playing two identical notes will result in the same *forces* being applied to the same mass. It's just that because the mass will be in motion as a result of the force applied by the first note, subsequent forces will result in different movements and therefore sounds. Thus, the effect of the performer actions are predictable - they always result in the same forces being applied to the physical model - but the virtual instrument response is not always the same. However, because musicians have experience of physical interactions in their everyday lives, the physical behaviour of the virtual instrument remains intuitively understandable.

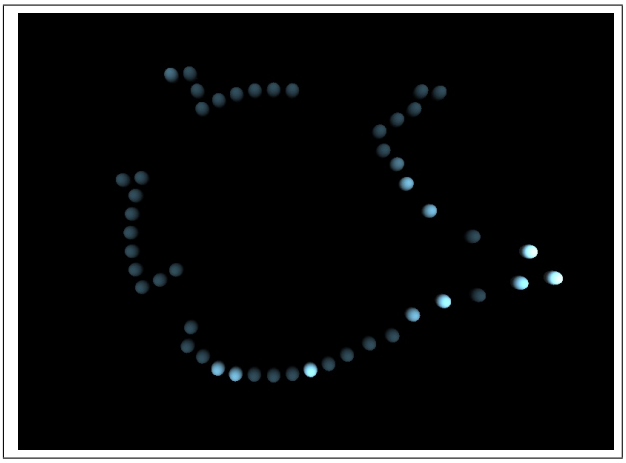

Figure 5. During performance the structure of the physical model was altered. This screenshot shows the model after a number of links have been cut and the tension in some springs relaxed.

In order to encourage a more conversational approach, at several points during performance the structure of the physical model was changed. The approximate points at which this would occur were pre-arranged with the musicians. The changes involved altering tension in some of the links between the masses and cutting others. The effect was that the circle would be seen to gradually lose shape as some of the masses broke loose (figure 5). This also resulted in a greater number of collisions between masses and thus a corresponding increase in more percussive sounds generated by the synthesis engine.

Altering the physical model during performance in this way was something we had not attempted previously. Our experience with *Partial Reflections 3* suggests that this is a technique which can help sustain conversational interaction over longer periods by allowing the virtual instrument to exhibit a wider range of behaviours. The challenge in future work will in developing techniques (musical and computational) for altering structures in this way while retaining transparency and providing sufficient support for instrumental interactions.

5. Conclusion

In this paper we have described our approach to virtual instrument design which involves using various techniques to facilitate what we call 'conversational' interaction. The concept of conversational interaction arose from a detailed study of the experiences of a small number of highly experienced professional musicians who used a series of virtual instruments we had designed for previous performances. Analysis of the data gathered during the studies indicated that the musicians demonstrated three 'modes of interaction' with the virtual instruments:

Instrumental In which the musician attempts to exert detailed control over all aspects of the virtual instrument.

Ornamental In which the musician does not attempt to actively alter the virtual instrument's behaviour or sound.

Conversational In which the musician shares control over the musical trajectory of the performance with the virtual instrument, seizing the initiative for a time to steer the conversation in a particular direction, then relinquishing control and allowing the virtual instrument to talk back.

We find conversational interaction the most interesting and challenging to design for and in this paper we have described several techniques that we used for a performance work called *Partial Reflections 3*. Specifically these techniques were:

- Using a simulated physical model to mediate between the live sounds produced on acoustic instruments and computer generated sounds and visuals. This underlying control structure helped facilitate conversational interaction because it could produce complex and occasionally surprising responses while retaining high-level controllability and transparency of operation.
- Enabling the musician to take an instrumental approach when desired by using consistent and intuitive mappings between the acoustic sounds and the state of the virtual instrument.
- Changing the structure of the physical model in relatively dramatic ways at several stages during performance.

A recording of a performance of *Partial Reflections 3* can be seen at http://www-staff.it.uts.edu.au/∼aj/videos/partial-reflections-III.mpg

6. Acknowledgments

The musicians Diana Springford and Jason Noble were co-creators of *Partial Reflections 3*. Our thanks to the musicians who participated in our user experience study for their time and insightful comments. Our thanks also to the developers of Pure Data, the Mass-Spring Damper objects and Graphical Environment for Multimedia for creating and making available the software which made this work possible. Finally, thankyou to the reviewers of this paper for providing very helpful feedback and suggestions.

This research was partly conducted within the Australasian CRC for Interaction Design (ACID), which is established and supported under the Australian Governments Cooperative Research Centres Program. An Australian Postgraduate Award provided additional financial support.

References

[1] A. Momeni and C. Henry, "Dynamic independent mapping layers for concurrent control of audio and video synthesis," *Computer Music Journal*, vol. 30, no. 1, pp. 49–66, 2006.

[2] I. Choi, "A manifold interface for kinesthetic notation in high-dimensional systems," in *Trends in Gestural Control of Music* (M. Wanderley and M. Battier, eds.), pp. 115–138, Paris: Ircam, 2000.

[3] J. Rovan, M. Wanderley, S. Dubnov, and P. Depalle, "Instrumental gestural mapping strategies as expressivity determinants in computer music performance," in *Proceedings of the AIMI International Workshop KANSEI - The Technology of Emotion, Genova*, pp. 68–73, 1997.

[4] A. Hunt, M. M. Wanderley, and R. Kirk, "Towards a model for instrumental mapping in expert musical interaction," in *Proc. International Computer Music Conference*, 2000.

[5] V. Krefeld and M. Waisvisz, "The hand in the web: An interview with Michel Waisvisz," *Computer Music Journal*, vol. 14, no. 2, pp. 28–33, 1990.

[6] B. Bongers, *Interactivation: Towards an e-cology of people, our technological environment, and the arts*. PhD thesis, Vrije Universiteit Amsterdam, 2006.

[7] J. Chadabe, "The limitations of mapping as a structural descriptive in electronic instruments," in *NIME '02: Proceedings of the 2002 conference on New interfaces for musical expression*, (Singapore, Singapore), pp. 1–5, National University of Singapore, 2002.

[8] P. Mogensen, "Towards a provotyping approach in systems development," *Scandinavian Journal of Information Systems*, vol. 4, pp. 31–53, 1992.

[9] A. Johnston, B. Marks, and L. Candy, "Sound controlled musical instruments based on physical models," in *Proceedings of the 2007 International Computer Music Conference*, pp. vol1: 232–239, 2007.

[10] B. G. Glaser and A. L. Strauss, *The discovery of grounded theory: strategies for qualitative research*. New York: Aldine de Gruyter, 1967.

[11] B. G. Glaser, *Theoretical Sensitivity*. The Sociology Press, 1978.

[12] A. Johnston, L. Candy, and E. Edmonds, "Designing and evaluating virtual musical instruments: facilitating conversational user interaction," *Design Studies*, vol. 29, no. 6, pp. 556–571, 2008.

[13] R. Rowe, *Interactive Music Systems*. The MIT Press, Cambridge, Mass., 1993.

[14] T. Winkler, *Composing Interactive Music: Techniques and Ideas Using Max*. Cambridge, MA, USA: MIT Press, 1998.

[15] M. S. Puckette, "Pure data," in *Proceedings of the International Computer Music Conference*, pp. 224–227, 1997.

[16] M. Danks, "The graphics environment for max," in *Proceedings of the International Computer Music Conference*, pp. 67–70, 1996.

[17] M. S. Puckette, T. Apel, and D. D. Zicarelli, "Real-time audio analysis tools for pd and msp," in *International Computer Music Conference*, (San Francisco), pp. 109–112, International Computer Music Association, 1998.

[18] C. Roads, *The Computer Music Tutorial*. Cambridge, MA, USA: MIT Press, 1996.